



Asamblea General

Distr. general
5 de junio de 2025
Español
Original: chino/español/francés/
inglés/ruso

Octogésimo período de sesiones

Tema 101 de la lista preliminar*
Desarme general y completo

La inteligencia artificial en el ámbito militar y sus implicaciones para la paz y la seguridad internacionales

Informe del Secretario General

Resumen

El presente informe ofrece un resumen consolidado de los principales aspectos de las comunicaciones que los Estados Miembros y los Estados observadores presentaron en respuesta a la resolución [79/239](#), sin perjuicio de sus posiciones respectivas. En sus secciones se presentan las oportunidades y retos que la inteligencia artificial plantea en el ámbito militar; una relación de las propuestas normativas existentes y emergentes; una recapitulación de las iniciativas sobre inteligencia artificial en el ámbito militar; algunas consideraciones sobre las próximas medidas; y las observaciones y conclusiones del Secretario General.

* [A/80/50](#).



Índice

	<i>Página</i>
I. Introducción	4
II. Antecedentes	4
III. Oportunidades y retos	4
IV. Propuestas normativas existentes y emergentes	8
V. Iniciativas de inteligencia artificial en el ámbito militar	10
VI. Próximas medidas	11
VII. Observaciones y conclusiones del Secretario General	13
 Anexo I	
Respuestas recibidas	15
A. Estados Miembros	15
Alemania	15
Argentina	19
Austria	20
Chile	23
China	26
Egipto	28
El Salvador	30
España	33
Federación de Rusia	36
Finlandia	41
Francia	43
Grecia	45
India	48
Indonesia	49
Irán (República Islámica del)	52
Israel	54
Italia	55
Japón	57
Lituania	61
México	62
Noruega	65
Nueva Zelanda	69
Países Bajos (Reino de los)	70
Pakistán	74

Perú	79
Reino Unido de Gran Bretaña e Irlanda del Norte	81
República de Corea	85
Serbia	89
Singapur	91
Suiza	93
Ucrania	98
B. Unión Europea	99

Anexo II

Replies received from international and regional organizations, the International Committee of the Red Cross, civil society, the scientific community and industry	101
A. International and regional organizations	101
African Commission on Human and Peoples' Rights	101
B. International Committee of the Red Cross	105
C. Civil society	109
Autonomists	109
Global Commission on Responsible Artificial Intelligence in the Military Domain ..	113
InterAgency Institute	118
International Committee for Robot Arms Control	120
International Humanitarian Law and Youth Initiative	121
Peace Movement Aotearoa and Stop Killer Robots Aotearoa New Zealand	125
Ploughshares	128
Soka Gakkai International	129
Stop Killer Robots	131
Stop Killer Robots Youth Network	134
Unione degli Scienziati Per Il Disarmo	138
Women's International League for Peace and Freedom	139
D. Scientific community	141
AI, Automated Systems, and Resort-to-Force Decision Making Research Project, the Australian National University	141
Queen Mary University of London, T.M.C. Asser Institute, University of Southern Denmark and University of Utrecht	146
United Nations Institute for Disarmament Research	150
E. Industry	155
Microsoft	155

I. Introducción

1. En el párrafo 7 de su resolución [79/239](#), relativa a la inteligencia artificial (AI) en el ámbito militar y sus implicaciones para la paz y la seguridad internacionales, la Asamblea General solicitó al Secretario General que recabase las opiniones de los Estados Miembros y de los Estados observadores sobre las oportunidades y los retos que planteaba para la paz y la seguridad internacionales la aplicación de la IA en el ámbito militar, con atención específica a los ámbitos distintos de los sistemas de armas autónomos letales, y que, en su octogésimo período de sesiones, le presentase un informe sustantivo en el que se resumieran esas opiniones y se clasificaran las propuestas normativas existentes y emergentes, con un anexo en el que figurasen esas opiniones, para que los Estados siguieran debatiendo. En el párrafo 8 de la misma resolución, la Asamblea solicitó también al Secretario General que solicitara las opiniones de las organizaciones internacionales y regionales, el Comité Internacional de la Cruz Roja, la sociedad civil, la comunidad científica y el sector industrial y las incluyera en el anexo del mencionado informe en el idioma original en que se recibieran. Este informe se presenta en atención a esas solicitudes.

2. El 12 de febrero de 2025, la Oficina de Asuntos de Desarme envió una nota verbal a todos los Estados Miembros y Estados observadores, en la que señalaba a su atención el párrafo 7 de la resolución [79/239](#) de la Asamblea General y recababa sus opiniones al respecto. También se enviaron notas verbales y cartas a las entidades especificadas en el párrafo 8 de la misma resolución, en las que se llamaba su atención sobre dicho párrafo y se recababan sus opiniones al respecto. Las opiniones recibidas al 11 de abril de 2025 se reproducen en los anexos del presente informe. Las opiniones que se reciban pasada esa fecha se publicarán en el sitio web de la Oficina en el idioma en que se presenten.

3. Las secciones II a VI del presente informe ofrecen un resumen consolidado de los principales aspectos de las comunicaciones recibidas de los Estados Miembros y los Estados observadores, sin perjuicio de sus posiciones respectivas. Las observaciones y conclusiones del Secretario General figuran en la sección VII.

II. Antecedentes

4. Los Estados se refirieron a los rápidos avances de la ciencia y la tecnología en general y de la IA en particular, indicando que estaban teniendo importantes repercusiones sociales. Concretamente, señalaron que la IA podía transformar todos los aspectos de los asuntos militares y tener importantes efectos en la paz y la seguridad internacionales.

5. Varios Estados hicieron referencia a aplicaciones actuales de la IA en el ámbito militar y a sus propios esfuerzos por utilizar la IA en operaciones de defensa. Aunque reconocían que era importante ocuparse de los sistemas de armas autónomos letales, los Estados observaron que la cuestión de la IA en el ámbito militar era más vasta y abarcaba un espectro más amplio de capacidades.

III. Oportunidades y retos

6. Se observó que la IA entrañaba oportunidades y retos que habían de abordarse de forma realista, aunque, según reconoció, en la actualidad no era posible preverlos todos debido a la rapidez a la que se estaba desarrollando esta tecnología y se recomendó que no se estigmatizara la propia tecnología.

A. Oportunidades

7. Se reconoció que la rapidez en el análisis de información y la adopción de decisiones era una de las principales ventajas de la IA. La escala fue otra de las ventajas señaladas, que permitía a la IA servir de “multiplicadora de fuerzas”. Varios Estados dijeron que la IA permitiría incrementar la eficiencia, la exactitud y la precisión y entrañaría menos probabilidades de error que el trabajo humano. La fiabilidad, la seguridad y la solidez fueron otras de las características apuntadas.

Aplicaciones

8. Varios Estados se refirieron a las aplicaciones de la IA en el ámbito de la inteligencia, la vigilancia y el reconocimiento, donde podía utilizarse para analizar con eficacia grandes series de datos, facilitar la detección de amenazas, mejorar la conciencia situacional y realizar operaciones más precisas. Se señaló que esas mismas características permitían a la IA respaldar la toma de decisiones y las funciones de mando y control, lo que permitiría afinar las operaciones, reducir los riesgos para la población civil y proteger en mayor medida los bienes de carácter civil. Con todo, también se hizo hincapié en que las herramientas de IA no debían suplantar la toma de decisiones humana.

9. Varios Estados indicaron que era posible integrar IA en sistemas no tripulados. Se apuntó que la IA podría mejorar la coordinación y la comunicación entre agentes militares y entre estos y otro tipo de agentes, como los proveedores de asistencia humanitaria. En términos generales, se observó que con la IA podría reducirse la carga de las tareas rutinarias o repetitivas y dedicar más capacidades humanas a tareas complejas.

10. Según algunos Estados, gracias a su capacidad de detectar intrusiones y otras actividades maliciosas, la IA podía utilizarse para mejorar la seguridad de las tecnologías de la información y las comunicaciones y proteger infraestructuras críticas. También podía servir para desenmascarar contenidos generados con esa misma tecnología con el propósito de difundir información errónea o desinformar y para detectar discursos de odio, propaganda o cambios de percepción pública.

11. Se mencionaron otros usos de IA que no estaban directamente relacionados con los combates, como la optimización logística, el mantenimiento predictivo, las adquisiciones, la asignación de recursos, la administración y las actividades de simulación y capacitación.

La paz y la seguridad internacionales

12. Varios Estados estimaban que la IA podía contribuir al mantenimiento de la paz y la seguridad internacionales; la conciencia situacional asistida por IA, por ejemplo, podía ayudar a mitigar riesgos y distender conflictos. Se señaló que con la IA podrían reducirse riesgos para el personal militar, por ejemplo, al sustituir a las personas en la realización de algunas tareas peligrosas, como la eliminación de municiones sin detonar, o respaldar operaciones de búsqueda y salvamento en lugares remotos.

13. Se apuntó que la IA podría mejorar la aplicación del derecho internacional humanitario, en particular sus principios fundamentales de distinción, proporcionalidad y precauciones en el ataque, así como la protección de los civiles y los bienes de carácter civil. A este respecto, varios Estados señalaron la capacidad de la IA para mejorar la conciencia situacional en general y la comprensión del entorno civil en particular, aumentar la precisión y reducir los riesgos de error humano. También se indicó que la IA podía facilitar las investigaciones sobre bajas civiles y hacer que los responsables rindieran cuentas.

14. Varios Estados apuntaron que la IA podría ayudar a supervisar y verificar la aplicación de acuerdos de desarme, no proliferación y control de armamentos. Se hizo referencia al potencial de apoyo a las misiones de mantenimiento de la paz que tenía esta tecnología, con la que, entre otras cosas, se facilitarían la planificación, la logística y la supervisión de acuerdos de alto el fuego. Se mencionaron otros ámbitos de aplicación de la IA, como la seguridad fronteriza, la lucha contra el terrorismo, la detección de programas ilegales de armamento y la optimización de la asistencia humanitaria y de la respuesta en casos de desastre.

B. Retos

15. Según varios Estados, la rápida evolución de las tecnologías emergentes en general y de la inteligencia artificial en particular planteaba retos para la paz y la seguridad internacionales que era importante comprender, aunque en la actualidad no era posible preverlos todos.

16. Se destacaron las siguientes preocupaciones en materia de IA:

- La aceleración del bucle “observar-orientar-decidir-actuar” está constriñendo los plazos de adopción de decisiones
- Su autonomía cada vez mayor y la pérdida de control humano, sobre todo en situaciones de uso de la fuerza
- Las posibilidades de utilizar esta tecnología de manera indebida o malintencionada
- El exceso de confianza de las personas en las aplicaciones de IA
- El aumento de las asimetrías tecnológicas entre Estados

La paz y la seguridad internacionales

17. Varios Estados señalaron que la integración de la IA en el ámbito militar podía plantear problemas para la paz y la seguridad internacionales. El riesgo de malentendidos, errores de cálculo e intensificación accidental de tensiones podría aumentar debido a la mayor rapidez y magnitud de las operaciones asistidas con IA o por fallos técnicos u otros motivos, lo que también podría hacer que disminuyese el umbral de uso de la fuerza. A algunos Estados les preocupaba que se desatase una carrera armamentista en este ámbito. Se planteó la posibilidad de que el uso de la IA llevara a pasar de las operaciones defensivas a las ofensivas y de que los crecientes desequilibrios entre Estados provocaran un aumento de la inestabilidad y acabaran por socavar la paz y la seguridad internacionales.

18. Varios Estados expresaron su preocupación por el efecto desestabilizador que podría generar la proliferación de capacidades de IA, también entre agentes no estatales. Se señaló que en la actualidad no había ningún marco multilateral para controlar la proliferación de armas con capacidades de IA.

Consideraciones tecnológicas

19. Los Estados plantearon diversos riesgos inherentes a la tecnología, como los siguientes:

- Fallos técnicos y averías
- Defectos de concepción
- Comportamientos indeseados que no responden a los parámetros de concepción

- Vulnerabilidad a ciberataques e intoxicaciones de datos
- Sesgos en los algoritmos y los datos, como los sesgos de género
- Sesgos de automatización resultantes de una capacitación deficiente de los operadores humanos
- Riesgos para la privacidad asociados a la recopilación y el procesamiento de grandes volúmenes de datos personales para adiestrar modelos de IA
- Problemas generados por modelos de IA mal adiestrados
- Problemas derivados de procedimientos deficientes de prueba, evaluación, validación y verificación
- Errores de selección de objetivos
- Consumo excesivo de energía
- Dependencia excesiva de proveedores externos

20. Hubo Estados que expresaron preocupación por la falta de transparencia y explicabilidad de dispositivos complejos de IA, a menudo calificados de “cajas negras”. También se plantearon preocupaciones por el uso de aplicaciones civiles, como la IA generativa, que podía introducir complejidad e incertidumbre en situaciones de conflicto, y por la convergencia de la IA con otras tecnologías.

Consideraciones humanitarias y de protección

21. Varios Estados observaron que la IA planteaba problemas de respeto del derecho internacional, en particular el derecho internacional humanitario y de los derechos humanos. El uso de la IA podía conducir a un uso indiscriminado de la fuerza y planteaba cuestiones de responsabilidad y rendición de cuentas en caso de actos ilegales o ilícitos. Se plantearon cuestiones conexas, como la protección de los civiles y las infraestructuras de carácter civil y la posibilidad de que se incrementase la intensidad y la letalidad de los conflictos para los combatientes.

22. Varios Estados plantearon preocupaciones éticas, observando que el uso de la IA podría dejar menos margen para la compasión, el razonamiento moral y el discernimiento humano.

Posibles ámbitos de uso indebido

23. Varios Estados plantearon la posibilidad de que la IA fuera utilizada por agentes estatales y no estatales para organizar ciberataques contra infraestructuras críticas u otros objetivos. También podía utilizarse para campañas de información engañosa y desinformación, para generar información falsa y ultrafalsificaciones y para propagar ese tipo de información mediante bots con IA. El uso de este tipo de desinformación y desinformación para influir en procesos electorales, por ejemplo, podría tener consecuencias desestabilizadoras.

Armas de destrucción masiva

24. Varios Estados subrayaron la importancia de que las armas nucleares y sus sistemas vectores se mantuvieran bajo control humano y expresaron su preocupación por la posibilidad de que la IA se integrase en los sistemas nucleares de mando, control y comunicación. Se hizo referencia al compromiso de determinados Estados poseedores de armas nucleares de mantener el control y la intervención humanos en todas las acciones críticas para informar y ejecutar decisiones soberanas relativas al empleo de armas nucleares. Se señalaron posibles repercusiones en la estabilidad estratégica y las tensiones.

25. Varios Estados expresaron preocupación por la posibilidad de que la IA facilitara la proliferación de armas de destrucción masiva destinadas, por ejemplo, a agentes no estatales. A este respecto, preocupaba en particular que esta tecnología pudiera ser utilizada para concebir y producir armas biológicas. Se señaló que, según las disposiciones de los tratados vigentes, la IA no debía utilizarse con ese fin. También se planteó que podía servir para contener la proliferación de armas de destrucción masiva.

IV. Propuestas normativas existentes y emergentes

26. Varios Estados dijeron que la IA debería utilizarse con fines pacíficos, como el arreglo pacífico de controversias. Otros subrayaron la importancia de abordar y mitigar los riesgos asociados la IA en el ámbito militar y, otros más, que esa tarea debía realizarse en colectivo.

27. Algunos Estados pidieron que la perspectiva desde la que se abordara la IA en el ámbito militar:

- Fuera flexible, equilibrada, realista y gradualista y permitiera así adaptarse a los avances tecnológicos
- Entrañara precauciones
- Abarcara todas las etapas del ciclo de vida de la IA, esto es, anteproyecto, proyecto, desarrollo, evaluación, prueba, despliegue, uso, venta, adquisición, utilización y retirada
- Se basara en las aplicaciones y usos de la IA, más que en la tecnología en sí misma
- Tuviera en cuenta las obligaciones contraídas

Se recomendó que las iniciativas orientadas en ese sentido distinguieran claramente entre usos letales y no letales.

Consideraciones jurídicas

28. Algunos Estados recordaron la resolución [79/239](#) de la Asamblea General, en la que se indicaba que el derecho internacional, incluidos la Carta de las Naciones Unidas, el derecho internacional humanitario y el derecho internacional de los derechos humanos, se aplicaba a las cuestiones que por él se regían y que surgían a lo largo de todas las etapas del ciclo de vida de la IA, incluidos los sistemas habilitados mediante la IA, en el ámbito militar. Se observó que el derecho internacional en general y el derecho internacional humanitario en particular no prohibían categóricamente el uso de capacidades de IA.

29. Los Estados sostuvieron que su uso de la IA en el ámbito militar se ajustaba al derecho internacional. Se afirmó que cumplir la ley y, en particular, las disposiciones del derecho internacional debía ser una consideración insoslayable en la gobernanza, la concepción y el despliegue de la IA en el ámbito militar. Se opinó asimismo que las aplicaciones de IA debían concebirse con el propósito de mejorar la aplicación del derecho internacional humanitario. Varios Estados subrayaron la importancia de controlar la legalidad de las nuevas armas y nuevos medios o métodos de guerra.

30. Varios Estados señalaron que, además de los marcos jurídicos, también importaban las consideraciones éticas.

Consideraciones sobre la paz y la seguridad internacionales

31. Varios Estados dijeron que la IA militar debía reforzar la paz y la seguridad internacionales y que sus aplicaciones no debían provocar inestabilidad o tensar situaciones. Se dijo que la IA no debía utilizarse para conseguir una supremacía militar absoluta y que los Estados debían velar por que esa tecnología no se convirtiera en una herramienta de invasión o en un medio de hegemonía.

32. Para algunos Estados, la IA no debía socavar los acuerdos vigentes de desarme, no proliferación y control de armamentos. Se instó a tratar de prevenir la proliferación de tecnologías de IA entre agentes no estatales y se subrayó la importancia de que no hubiera mecanismos internacionales de supervisión arbitrarios ni controles de exportación discriminatorios.

Uso responsable de la inteligencia artificial en el ámbito militar

33. En opinión de algunos Estados, la IA debía aplicarse de forma responsable durante todo su ciclo de vida. Se dijo que el concepto de responsabilidad debía implicar legalidad y rendición de cuentas.

34. Varios Estados subrayaron la importancia de abordar la IA con un enfoque centrado en el ser humano. Muchos consideraban que esta tecnología debía estar sujeta en todo momento a control y responsabilidad humanos. Se hizo referencia a conceptos como “control y juicio humanos en función del contexto” y “control humano determinante”, importantes para unos pero mal definidos para otros. Se dijo que la noción de “control humano determinante” podía utilizarse para entorpecer actividades legítimas de investigación o restringir indebidamente el uso de la IA en el ámbito militar.

35. A algunos Estados les parecía importante que siguiera habiendo personas que asumieran la responsabilidad y rindieran cuentas, entre otras cosas estableciendo una cadena responsable de mando y control humanos, de conformidad con el derecho internacional.

Consideraciones tecnológicas

36. Los Estados consideraron los siguientes principios para la gobernanza de tecnologías:

- Seguridad, para garantizar que los sistemas de IA resisten amenazas externas
- Protección, mediante la incorporación de salvaguardas que permitan reducir al máximo los daños
- Fiabilidad, para evitar imprevistas y defectos de funcionamiento
- Límites y restricciones operacionales claros para evitar comportamientos no deseados
- Usos bien definidos
- Gobernabilidad, gracias a una interacción humano-máquina adecuada y a la mitigación de los sesgos
- Equidad e imparcialidad
- Protección de la privacidad
- Explicabilidad, inteligibilidad y trazabilidad
- Transparencia

37. Se insistió en que en las actividades de capacitación se utilizaran datos que permitieran cumplir plenamente el derecho internacional. Varios Estados subrayaron la importancia de hacer controles durante todo el ciclo de vida para detectar fallos en los sistemas y asegurarse de que son fiables. También se subrayó la necesidad de capacitar adecuadamente al personal que trabaja con IA para mitigar los riesgos y garantizar que se respete el derecho internacional humanitario y de supervisar el funcionamiento de los sistemas a lo largo de su ciclo de vida y cuidar de que se desactiven de forma segura cuando se retiren.

V. Iniciativas de inteligencia artificial en el ámbito militar

Foros internacionales

38. Varios Estados hicieron referencia a debates que estaban desarrollándose en las Naciones Unidas, al Pacto para el Futuro (resolución 79/1 de la Asamblea General) y al Pacto Digital Global que lleva anexo, así como a la resolución de la Asamblea General sobre la inteligencia artificial en el ámbito militar y sus implicaciones para la paz y la seguridad internacionales (resolución 79/239). También se mencionó la reunión del Consejo de Seguridad sobre la utilización de una IA segura, inclusiva y fiable para el mantenimiento de la paz y la seguridad internacionales, celebrada con arreglo a la fórmula Arria el 4 de abril de 2025.

39. Algunos Estados hablaron de los debates multilaterales sobre cuestiones relativas al uso militar de la IA, como las deliberaciones de la Comisión de Desarme en torno al tema del programa titulado “Recomendaciones sobre los entendimientos comunes relativos a las tecnologías emergentes en el contexto de la seguridad internacional”, a la labor del Grupo de Expertos Gubernamentales sobre las Tecnologías Emergentes en el Ámbito de los Sistemas de Armas Autónomos Letales, creado en el marco de la Convención sobre Prohibiciones o Restricciones del Empleo de Ciertas Armas Convencionales que Puedan Considerarse Excesivamente Nocivas o de Efectos Indiscriminados, y a las resoluciones de la Asamblea General sobre los sistemas de armas autónomos letales (resoluciones 78/241 y 79/62).

40. Los Estados también se refirieron a su participación en actividades relacionadas con la aplicación militar de la IA organizadas por la Oficina de Asuntos de Desarme y el Instituto de las Naciones Unidas de Investigación sobre el Desarme.

Iniciativas de Estados

41. Varios Estados indicaron que habían puesto en marcha o participado en iniciativas relacionadas con la aplicación militar de la IA, como las siguientes:

- La Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar y su consiguiente proceso de aplicación
- El proceso sobre Inteligencia Artificial Responsable en el Ámbito Militar, que comprendió dos cumbres, una celebrada en el Reino de los Países Bajos en 2023, en la que se aprobó un Llamamiento a la Acción, y otra celebrada en la República de Corea en 2024, en la que se aprobó un Plan de Acción. Se espera que una Comisión Global sobre Inteligencia Artificial Responsable en el Ámbito Militar publique un informe al respecto antes de la próxima cumbre, que se celebrará en España en 2025
- La Cumbre de Acción sobre Inteligencia Artificial celebrada en Francia en 2025, en la que se aprobó la Declaración de París sobre el Mantenimiento del Control Humano en los Sistemas de Armas Dotados de IA

- La Cumbre sobre Seguridad de la IA que se celebró en el Reino Unido de Gran Bretaña e Irlanda del Norte en 2023, en la que se aprobó la Declaración de Bletchley
- Los trabajos sobre IA desarrollados en el marco del Grupo de los Siete
- La alianza por la IA en el ámbito de la Defensa
- La Iniciativa Mundial de Gobernanza de la IA presentada en 2023

42. Se dijo que esas iniciativas eran útiles, pero podían generar fragmentación. También preocupaba que sus resultados no tuvieran en cuenta las opiniones de todos los Estados implicados y pudieran socavar esfuerzos de inclusión en este ámbito.

Iniciativas regionales

43. Algunos Estados observaron que las iniciativas regionales eran importantes para que los debates sobre la IA en el ámbito militar fueran inclusivos y abordaran circunstancias concretas. He aquí algunos ejemplos:

- La declaración conjunta sobre cooperación en la aplicación de la IA en el sector de la defensa, aprobada durante la Estancia de Encuentro de Ministros de Defensa de la Asociación de Naciones de Asia Sudoriental de la (ASEAN) en 2025
- La 16^a Conferencia de Ministros de Defensa de las Américas, celebrada en 2024, en la que se aprobó la Declaración de Mendoza
- Las actividades desarrolladas en el marco de la Organización del Tratado del Atlántico Norte, como la estrategia de IA de esta entidad, revisada por última vez en 2024, y los principios de uso responsable que presentó en 2021
- Consultas regionales en el marco del proceso de Inteligencia Artificial Responsable en el Ámbito Militar que se realizaron en 2024 en Chile, Kenya, Países Bajos (Reino de los), Singapur y Türkiye

Iniciativas nacionales

44. Hubo Estados que hablaron de herramientas nacionales, como las disposiciones legislativas, los reglamentos, las estrategias y los organismos que ya se ocupaban de la IA, y de los esfuerzos desplegados para crearlas.

VI. Próximas medidas

45. Los Estados pidieron un diálogo sobre la IA en el ámbito militar. Varios de ellos pidieron que se siguieran estudiando las implicaciones del uso militar de la IA para la paz y la seguridad internacionales.

46. Numerosos Estados indicaron que el propósito de seguir dialogando debería ser mitigar los riesgos que plantea la IA en el ámbito militar. Se propuso que ese diálogo se orientara a la elaboración de marcos normativos o de gobernanza. Varios Estados instaron a establecer normas, reglas y principios que regulen todo el ciclo de vida de las aplicaciones militares de la IA. Algunos preconizaban el establecimiento de un marco jurídicamente vinculante, pero a otros no les parecía necesario aprobar nuevas disposiciones jurídicas en esos momentos. También se dijo que el establecimiento de normas, reglas y principios jurídicamente vinculantes podría acordarse ulteriormente. Varios Estados se oponían a establecer normas, reglas y principios de desarrollo, despliegue o utilización responsables, aduciendo que la idea no gozaba de consenso. Se recomendó no establecer normas prematuramente.

47. Se insistió en la importancia de no fragmentar la gobernanza y evitar duplicidades en ese ámbito. Se consideraba que al abordar la gobernanza debían sopesarse debidamente las cuestiones humanitarias, de seguridad y de desarrollo. Algunos Estados subrayaron la importancia de evitar restricciones que entorpecieran los procesos legítimos de innovación y progreso tecnológico. Varios pensaban que las aplicaciones pacíficas de la IA, sobre todo las de países en desarrollo, no debían restringirse.

48. Se recomendó tener en cuenta en toda estrategia de gobernanza que los Estados se encuentran en diferentes estadios de integración militar de la IA y tienen condiciones de seguridad diversas. Se destacó la importancia de que todos los Estados participaran en los debates sobre la gobernanza de la IA en el ámbito militar. Muchos Estados estimaban que los futuros debates debían tener un carácter multipartito y entrañar la participación de organizaciones internacionales y regionales, la sociedad civil, la comunidad científica y las empresas del sector. Se afirmó que, en cambio, la toma de decisiones debía seguir siendo prerrogativa exclusiva de los Estados.

49. Los Estados consideraban que el futuro diálogo sobre el uso militar de la IA debería articularse en torno a las prioridades siguientes:

- Asegurarse de que se respeta el derecho internacional, en particular el derecho internacional humanitario
- Proteger la dignidad y los derechos humanos
- Tratar de concertar términos y definiciones
- Considerar la posibilidad de aplicar medidas de transparencia y fomento de la confianza
- Abordar la cuestión de la autonomía en el uso de la fuerza
- Tratar de los sistemas de IA que se utilizan directamente en operaciones de combate
- Asegurarse de que haya mecanismos adecuados de gobernanza de los datos
- Reforzar las iniciativas internacionales de cooperación y asistencia
- Contribuir al desarrollo de capacidades mediante el intercambio de conocimientos y buenas prácticas, la transferencia de tecnologías y por otros medios para ir cerrando la brecha digital y las distancias en materia de IA
- Promover un diálogo regional permanente
- Promover normas nacionales que, entre otras cosas, garanticen la observancia del derecho internacional en el sector privado

50. Varios Estados propusieron que la cuestión de los sistemas de armas autónomos letales se tratara en todos los debates sobre el uso militar de la IA. También se dijo que los debates que se estaban desarrollando en relación con esa tecnología y con el uso militar de la IA eran complementarios. Diversos Estados recordaron sus posiciones en lo tocante a los sistemas de armas autónomos letales¹. Para algunos Estados, el Grupo de Expertos Gubernamentales establecido en el marco de la Convención sobre Ciertas Armas Convencionales era el foro idóneo para abordar la IA en el ámbito militar, pero otros no lo consideraban adecuado debido a su mandato específico y a que su composición no era universal.

51. Varios Estados pidieron que el uso militar de la IA se abordara en los foros de las Naciones Unidas. Se propuso que utilizaran el presente informe como base de esos

¹ Para más información, véase A/79/88.

debates. Según algunos Estados, los futuros debates deberían complementar los procesos que ya están desarrollándose, como el grupo de trabajo de composición abierta sobre la seguridad de las tecnologías de la información y las comunicaciones y de su uso.

52. Varios Estados estimaban que el mecanismo de desarme de las Naciones Unidas era una plataforma inclusiva eficaz que debía desempeñar un papel central en los futuros debates sobre la IA en el ámbito militar. Se propuso que la Conferencia de Desarme se ocupara de la IA, sobre todo en relación con las armas nucleares. Se dijo que la Primera Comisión de la Asamblea General también podía tratar el tema y encargar al Secretario General informes periódicos sobre el estado de desarrollo tecnológico de la IA en el ámbito militar. Varios Estados plantearon la posibilidad de que el asunto se debatiera en el seno de la Comisión de Desarme.

53. También se planteó la posibilidad de someter la cuestión a la deliberación del Consejo de Seguridad.

54. Varios Estados preconizaban establecer un grupo de trabajo de composición abierta o algún otro proceso específico. A otros les parecía que en ese momento no procedía iniciar un nuevo proceso en las Naciones Unidas. También se dijo que cualquier proceso de las Naciones Unidas sobre ese tema debía ser objeto de consenso.

VII. Observaciones y conclusiones del Secretario General

55. La IA podría repercutir en todas las facetas de nuestras vidas. Cuando se utiliza para fines pacíficos, puede facilitar considerablemente el cumplimiento de compromisos y objetivos de desarrollo, como los Objetivos de Desarrollo Sostenible.

56. En el ámbito militar, la IA puede beneficiar tanto a los ejércitos que la emplean como a las poblaciones civiles, ya que permite mejorar la precisión de las operaciones y reducir el margen de error humano. Pero aplicar la IA en el ámbito militar plantea también grandes retos y el más importante de ellos es mantener una asunción de responsabilidades y una rendición de cuentas humanas.

57. La afirmación que la Asamblea General hace en su resolución [79/239](#) de que el derecho internacional, incluidos la Carta de las Naciones Unidas, el derecho internacional humanitario y el derecho internacional de los derechos humanos, se aplica a lo largo del ciclo de vida de las capacidades de IA, es una importante base de partida, pero aún falta por resolver algunas cuestiones importantes sobre la aplicación de esos instrumentos.

58. Ha de prestarse particular atención al empleo de capacidades militares de IA en situaciones de uso de la fuerza. Aunque la IA podría utilizarse para mejorar la protección de civiles y combatientes, los usos conocidos de esta tecnología en conflictos actuales plantean problemas de control humano y pueden dar pábulo a las hostilidades en zonas densamente pobladas. Sería políticamente inaceptable y moralmente repugnante que hubiera máquinas con capacidad discrecional de acabar con vidas humanas.

59. Los riesgos que plantean las armas nucleares no desaparecerán mientras no se eliminen las propias armas. A la espera de la eliminación total de las armas nucleares, insto a todos los Estados que poseen armas de ese tipo a acordar que las decisiones relativas a su uso las tomen personas, no máquinas.

60. La IA puede facilitar a agentes estatales y no estatales posibilidades de desarrollar o adquirir armas químicas y biológicas. Insto pues a los Estados a cumplir plenamente las obligaciones a que están sujetos en virtud de los instrumentos

pertinentes de desarme, no proliferación y control de armamentos evaluando sistemáticamente los riesgos que entraña la IA y sus consecuencias para esos instrumentos y preparándose adecuadamente para afrontarlos.

61. La introducción de aplicaciones civiles de la IA en el ámbito militar es un motivo de creciente preocupación. El hecho de que las tecnologías de IA sean por naturaleza reutilizables plantea problemas de supervisión, transparencia y rendición de cuentas. Insto a los Estados a que examinen detenidamente la rapidez cada vez mayor con la que las nuevas aplicaciones civiles de la IA se adoptan en el ámbito militar.

62. La búsqueda de otros mecanismos de cooperación en materia de IA, sobre todo de ámbito regional y subregional, resulta muy valiosa. Las organizaciones regionales y subregionales son las más indicadas para concebir y poner en práctica medidas de transparencia y fomento de la confianza que permitan mitigar los riesgos. Por consiguiente, aliento a los Estados a que consideren la posibilidad de concebir medidas regionales y subregionales de transparencia y fomento de la confianza atendiendo a las singulares características de la IA y a los retos que plantea.

63. En las Naciones Unidas se están promoviendo debates inclusivos sobre los usos pacíficos de la IA y sobre cómo gobernar esta tecnología en beneficio de la humanidad, particularmente en el contexto de la aplicación del Pacto Digital Global. Ahora bien, los Estados Miembros han estado analizando la aplicación militar de esta tecnología sobre todo fuera de los foros de las Naciones Unidas. La resolución [79/239](#) de la Asamblea General y el presente informe son dos primeros pasos importantes para traer a las Naciones Unidas este debate trascendental. Aliento a los Estados a mantener unas deliberaciones inclusivas y constructivas que permitan avanzar en el establecimiento de entendimientos compartidos y reforzar la cooperación internacional para mitigar los riesgos.

64. Se alienta a los Estados a que estudien medidas de creación de capacidad o de otra índole para asegurarse de que todos los Estados participan de manera efectiva en los procesos de las Naciones Unidas sobre este tema, lo que es indispensable para promover un punto de vista compartido, desarrollar planteamientos comunes y mitigar posibles riesgos.

65. La Asamblea General ha mostrado gran habilidad al iniciar procesos que fomentan debates inclusivos sobre cuestiones relacionadas con las tecnologías emergentes y la seguridad internacional y facilitan al mismo tiempo la implicación de organizaciones internacionales y regionales, la sociedad civil, la comunidad científica, las empresas del sector y demás interesados. Este enfoque multipartito es especialmente importante en el ámbito de la IA, en el que la innovación está en gran medida impulsada por el sector privado y el grueso del conocimiento técnico no está en manos de Gobiernos, sino de entidades académicas y científicas.

66. Recomiendo que los Estados estudien las ideas planteadas en este informe y que, en el octogésimo período de sesiones de la Asamblea General, adopten medidas concretas para establecer un proceso específico de carácter inclusivo que permita abordar en su globalidad la aplicación militar de la IA y sus implicaciones para la paz y la seguridad internacionales.

Anexo I

Respuestas recibidas

A. Estados Miembros

Alemania

[Original: inglés]
[11 de abril de 2025]

I. Introducción

En los últimos años se ha producido una evolución sin precedentes en las tecnologías de inteligencia artificial (IA) que ha entrañado el desarrollo de aplicaciones basadas en tecnologías disruptivas, como la IA generativa. Es indispensable que los Estados puedan aprovechar las oportunidades derivadas de estos avances y asegurarse de que no se entorpece el progreso tecnológico. Al mismo tiempo, los Estados deben garantizar que las aplicaciones de la IA en el ámbito militar se desarrollem y utilicen de forma responsable y respetando plenamente el derecho internacional, incluido el derecho internacional humanitario. El diálogo internacional es indispensable para lograr este difícil equilibrio.

En este contexto, Alemania contribuye activamente a los procesos internacionales relativos al uso responsable de la IA en el ámbito militar. Formó parte del grupo básico de países que copatrocinaron la resolución [79/239](#) de la Asamblea General, relativa a la inteligencia artificial en el ámbito militar y sus implicaciones para la paz y la seguridad internacionales, y apoya plenamente los esfuerzos desplegados por el Secretario General para presentar un informe sustantivo con las opiniones de los Estados Miembros sobre “las oportunidades y los retos que plantea para la paz y la seguridad internacionales la aplicación de la inteligencia artificial en el ámbito militar”.

Alemania celebra la oportunidad de examinar más a fondo las opiniones de los Estados Miembros y otras partes interesadas y de presentar algunos aspectos de sus propias consideraciones sobre estas importantes cuestiones.

II. Principios e hipótesis de trabajo

La estrategia de Alemania Para garantizar el uso militar responsable de la IA se basa en los principios fundamentales, definidos en diversos foros y debates internacionales, que se indican a continuación.

Alemania contribuyó activamente a formular en 2021 los principios de uso responsable de la Organización del Tratado del Atlántico Norte (OTAN), con los que sigue estando plenamente conforme: la legalidad en el desarrollo y el uso de aplicaciones de inteligencia artificial; la responsabilidad humana de garantizar que exista rendición de cuentas en el diseño y el funcionamiento de la inteligencia artificial en los sistemas militares; la explicabilidad y trazabilidad de las aplicaciones de inteligencia artificial en el ámbito militar; el mantenimiento de la fiabilidad, la seguridad y la solidez de los sistemas dotados de inteligencia artificial y autonomía durante todo su ciclo de vida; y la gobernabilidad mediante la interacción adecuada entre el ser humano y la máquina y la mitigación de los sesgos.

Además, Alemania aprobó los documentos finales de las dos ediciones de la Cumbre sobre la Inteligencia Artificial Responsable en el Ámbito Militar, que se

celebraron en La Haya en 2023 (Llamamiento a la Acción) y en Seúl en 2024 (Plan de Acción), así como la Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar promovida por los Estados Unidos de América en 2023, en cuya aplicación participa activamente.

También forma parte de la iniciativa AI Partnership for Defence, con la que países de ideas afines fomentan el uso responsable de la IA, promueven intereses comunes y mejores prácticas de aplicación ética de la IA, establecen marcos para facilitar la cooperación y coordinan mensajes estratégicos sobre políticas de IA.

En febrero de 2025, Alemania aprobó la Declaración de París sobre el mantenimiento del control humano en los sistemas de armamentos que funcionan con IA, que pone de relieve la importancia de mantener bajo control humano la aplicación de la IA en el ámbito militar.

III. Principales aspectos del uso de la inteligencia artificial en las Fuerzas Armadas Federales de Alemania

Las Fuerzas Armadas Federales de Alemania (Bundeswehr) están considerando la posibilidad de utilizar la IA para cumplir su misión principal y ganar superioridad en materia de información, toma de decisiones y eficacia, y para optimizar los procesos administrativos, logísticos y de mantenimiento predictivo de sistemas complejos. La IA también se utiliza para asistir al personal especializado en tareas de carácter civil-militar de detección temprana de crisis valiéndose de diversas competencias de análisis masivos de datos y para hacer proyecciones de despliegues. La IA es un componente esencial de grandes proyectos de defensa, que también se llevan a cabo en un contexto europeo, lo que contribuye a mantener y fomentar la excelencia tecnológica europea. Por lo que respecta al desarrollo nacional y tecnológico en el sector internacional del armamento, la IA sirve para garantizar que en el futuro se tengan las capacidades necesarias de defensa nacional y aliada. El establecimiento de medios de despliegue de la IA, en particular para fines militares y de seguridad nacional, se lleva a cabo dentro de las atribuciones y responsabilidades de los ministerios y departamentos competentes. Sin perjuicio de lo anterior, las tecnologías y aplicaciones de IA utilizadas en el ámbito de la seguridad se recogen en la estrategia de IA del Gobierno Federal del país.

La Bundeswehr, que tiene la máxima exigencia ética, establece las normas jurídicas más estrictas sobre el uso de la IA en los sistemas de armamentos. Concretamente, sigue las disposiciones del derecho internacional humanitario aplicables a los conflictos armados y las directrices de la Comisión de Ética de Datos del Gobierno Federal y de la OTAN, en especial los seis principios de uso responsable antes mencionados, para el uso militar de la IA durante todo el ciclo de vida de esos sistemas.

IV. Consideraciones esenciales

Para mantener las capacidades de defensa y disuasión necesarias, Alemania sigue decidida a aprovechar las oportunidades que ofrece la IA en el ámbito militar y está convencida de que no se debe entorpecer el progreso tecnológico, sobre todo en vista del doble uso inherente a las tecnologías en cuestión.

Al mismo tiempo, Alemania seguirá ampliando la base de conocimientos mediante la evaluación y el tratamiento de los riesgos asociados al uso de la IA en el ámbito militar, por ejemplo los sesgos involuntarios, como los relativos al género. Alemania concede gran importancia al papel esencial del mundo académico y a las

valiosas aportaciones de institutos de investigación y grupos de reflexión que trabajan en este ámbito. Para fomentar la investigación sobre este tema, Alemania presta apoyo a organizaciones de investigación competentes, como el Instituto de las Naciones Unidas de Investigación sobre el Desarme (UNIDIR), y aporta financiación a proyectos de investigación centrados en objetivos concretos.

Para Alemania, es de suma importancia que los debates sean inclusivos desde el punto de vista geográfico y tengan en cuenta las opiniones no solo de los Estados Miembros, sino también del sector privado, la sociedad civil y el mundo académico.

Al abordar las oportunidades y los riesgos asociados a los sistemas de armas que funcionan con IA, Alemania concede especial importancia al concepto de control humano y considera que disponer de un marco eficaz de control humano es una condición necesaria para garantizar que todos los sistemas de armas se atengán al derecho internacional humanitario, lo que, además del control técnico, requiere un elemento de juicio. El concepto alemán de marco de control humano engloba un conjunto de actuaciones y medidas tecnológicamente posibles que establecen unos límites claros dentro de los cuales se permite operar al algoritmo del sistema. El derecho internacional, y en particular el derecho internacional humanitario, es un elemento central de esos límites. Cuando se trata del uso efectivo de la IA en el campo de batalla, el contexto es de suma importancia. Alemania considera que el concepto de marco de control humano es una forma adecuada de tenerlo debidamente en cuenta.

Debe prestarse especial atención al uso de la IA en armas nucleares, una cuestión en la que el debate científico y político todavía está en ciernes. El posible uso de la IA en los sistemas de mando y control de las armas nucleares podría tener graves repercusiones para la estabilidad estratégica o la escalada nuclear. Al mismo tiempo, la IA podría abrir nuevas posibilidades de contener la propagación y el uso de armas de destrucción masiva. Alemania trató de contribuir a esos debates organizando una conferencia sobre inteligencia artificial y armas de destrucción masiva en el marco de su consolidada serie de conferencias sobre el tema “Asimilar la tecnología - replantear el control de armas”, que se celebrará en Berlín el 28 de junio de 2024.

La Convención sobre la Prohibición del Desarrollo, la Producción y el Almacenamiento de Armas Bacteriológicas (Biológicas) y Toxínicas y sobre Su Destrucción y la Convención sobre la Prohibición del Desarrollo, la Producción, el Almacenamiento y el Empleo de Armas Químicas y sobre Su Destrucción prohíben categorías enteras de armas de destrucción masiva. Aplicaciones como los grandes modelos lingüísticos (generativos) pueden facilitar la proliferación de conocimientos de doble uso que podrían emplearse indebidamente para desarrollar, producir o utilizar armas biológicas y químicas. La convergencia de aplicaciones de IA, como AlphaFold, con la biología sintética puede permitir que agentes malintencionados diseñen proteínas novedosas que, al tener modificaciones en la secuencia del ADN, no sean detectadas. La IA puede utilizarse para analizar cúmulos de macrodatos, como los correspondientes al genoma humano, y ser de gran provecho para el desarrollo de terapias médicas individuales, pero también podría usarse indebidamente para desarrollar armas biológicas dirigidas contra grupos étnicos específicos.

Por ello, en estrecha colaboración con sus asociados internacionales, Alemania seguirá estudiando posibles líneas de actuación a fin de evaluar la incidencia de las aplicaciones de la IA en el desarrollo y la producción de armas prohibidas y establecer posibles regulaciones. Al mismo tiempo, Alemania aprovechará las ventajas de la IA en los ámbitos de la verificación, la biología forense y la reducción de riesgos.

V. El compromiso de Alemania con los procesos internacionales

Alemania ha contribuido activamente desde sus inicios al proceso de la Inteligencia Artificial Responsable en el Ámbito Militar y seguirá haciéndolo. Formó parte del grupo básico de copatrocinadores de la resolución de la Asamblea General 79/239, relativa a la inteligencia artificial en el ámbito militar y sus implicaciones para la paz y la seguridad internacionales. Alemania encomia encarecidamente el enfoque interregional y multipartito de esta importante iniciativa y espera con interés su continuación en España en septiembre de 2025.

De forma totalmente complementaria, Alemania ha contribuido a la Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar promovida por los Estados Unidos, por ejemplo al copresidir el grupo de trabajo sobre supervisión (junto con Austria).

También participa activamente en la iniciativa AI Partnership for Defence y en la red de expertos del UNIDIR sobre gobernanza de la inteligencia artificial en el ámbito militar.

Alemania apoya al Presidente del Grupo de Expertos Gubernamentales sobre las Tecnologías Emergentes en el Ámbito de los Sistemas de Armas Autónomos Letales con sede en Ginebra, el Embajador Robert in den Bosch , y sigue contribuyendo de forma activa al proceso, entre otras cosas coordinando las posiciones de varios Estados Miembros en el marco del denominado grupo de dos niveles. En estrecha colaboración con sus asociados internacionales, Alemania seguirá trabajando para que el Grupo culmine oportunamente su mandato, preferiblemente para finales de 2025.

En el contexto de la OTAN, Alemania reconoce el potencial de la IA para el desarrollo de las fuerzas armadas y la capacidad de defensa de la alianza, así como los retos que entraña su uso para la interoperabilidad de las fuerzas armadas de los países miembros. Deben tenerse plenamente en cuenta los desarrollos multinacionales de la IA y las medidas de normalización de esta tecnología en el seno de la OTAN, la Unión Europea y los países asociados de Alemania, para garantizar la interoperabilidad de la fuerza militar de la Bundeswehr en operaciones internacionales. Por ello, Alemania acogió con satisfacción que, en la estrategia de IA de la OTAN, los países miembros acordaran unos principios de uso responsable.

VI. Perspectivas futuras

Como las tecnologías disruptivas emergentes seguirán evolucionando y conformando nuestro mundo, Alemania considera indispensable establecer una coordinación internacional inclusiva del desarrollo y el uso militar responsable de la IA. Los procesos internacionales existentes ofrecen un excelente marco para abordar los aspectos sustanciales y considerar las opiniones de diversas partes interesadas pertinentes. Alemania seguirá contribuyendo activamente a estos esfuerzos para poner en práctica compromisos políticos sobre el uso militar responsable de la IA, como la Declaración Política promovida por los Estados Unidos y el proceso de Inteligencia Artificial Responsable en el Ámbito Militar, y ampliar el apoyo que reciben. Alemania está deseosa de examinar los resultados del informe del Secretario General sobre la IA en el ámbito militar. Alemania seguirá contribuyendo activamente al proceso dedicado a los sistemas de armas autónomos letales en el marco del Grupo de Expertos Gubernamentales de Ginebra.

Argentina

[Original: español]
 [10 de abril de 2025]

Con relación a la resolución [79/239](#) titulada “La inteligencia artificial en el ámbito militar y sus implicaciones para la paz y la seguridad internacionales”, aprobada por la Asamblea General el 24 de diciembre de 2024, se remite a continuación el informe correspondiente.

Enfoque general

La República Argentina reconoce que la evolución de la inteligencia artificial (IA) en el entorno militar presenta impactos estratégicos significativos. Su aplicación genera beneficios concretos en diversas funciones no letales, al tiempo que introduce riesgos que requieren atención desde el derecho internacional, la ética y la responsabilidad operativa. En ese marco, el desarrollo y uso de estas tecnologías debe mantenerse bajo el respeto al derecho internacional humanitario y los derechos humanos, garantizando en todo momento la responsabilidad y el control humanos en la toma de decisiones críticas.

Oportunidades

La IA en defensa, particularmente en sus aplicaciones no letales, constituye una herramienta legítima y valiosa para la mejora de capacidades nacionales. Entre los usos prioritarios se destacan los siguientes:

- Optimización logística y operativa
- Apoyo al procesamiento de inteligencia
- Fortalecimiento de la ciberdefensa
- Simulación, entrenamiento y planificación estratégica

Estas capacidades contribuyen a operaciones más eficientes, seguras y adaptadas a escenarios contemporáneos, reforzando la eficacia defensiva sin comprometer los principios humanitarios ni las obligaciones internacionales del Estado.

Desafíos

El desarrollo acelerado de la IA en contextos militares plantea desafíos que deben ser abordados colectivamente, entre los cuales se destacan los siguientes:

- La reducción del umbral para el uso de la fuerza y el acortamiento del tiempo de decisión humana
- La posibilidad de sesgos algorítmicos no detectados
- La proliferación de sistemas autónomos a actores no estatales
- El riesgo de consolidación de asimetrías tecnológicas entre Estados
- Estos riesgos refuerzan la necesidad de establecer principios comunes, salvaguardas verificables y marcos cooperativos

Gobernanza, cooperación internacional e inclusión tecnológica

Entendemos que cualquier proceso normativo en esta materia debe construirse sobre los siguientes principios:

- Evitar regulaciones generales o prematuras que limiten el desarrollo autónomo de tecnologías defensivas legítimas
- Distinguir claramente entre usos letales y no letales
- Garantizar el control humano significativo como condición operativa y normativa ineludible
- Promover una cooperación internacional inclusiva, orientada al fortalecimiento de capacidades y a la reducción de brechas tecnológicas entre Estados

La Argentina ha reafirmado estos principios en instancias multilaterales recientes, destacando la importancia de avanzar en normas compartidas para el uso responsable de la IA en el ámbito militar, particularmente en relación con ciberdefensa y ciberseguridad.

A modo de ejemplo como iniciativa a nivel regional, se destaca la realización, en el marco de la XVI Conferencia de Ministros de Defensa de las Américas 2024 llevada a cabo en Mendoza, Argentina, de la reunión de la comisión de trabajo sobre el desarrollo responsable, la aplicación y la gobernanza de la inteligencia artificial en el ámbito militar, con el objetivo de trabajar en conjunto el desarrollo de estándares internacionales.

Sobre la referencia al Pacto para el Futuro

Finalmente, se deja constancia de que la República Argentina se ha disociado formalmente del Pacto para el Futuro, citado en el preámbulo de la resolución [79/239](#) de la Asamblea General. Dicha mención no representa, por tanto, compromiso, adhesión ni respaldo por parte del Estado argentino.

Austria

[Original: inglés]
[11 de abril de 2025]

En respuesta a lo solicitado en el párrafo 7 de la resolución [79/239](#) de la Asamblea General, Austria desea compartir las siguientes reflexiones y observaciones nacionales.

Inteligencia artificial para la ciberseguridad y la ciberdefensa

Ya se utilizan ampliamente programas informáticos de ciberseguridad que funcionan con inteligencia artificial (IA) para ayudar a detectar intrusiones y demás actividades maliciosas en redes de computadoras. Es probable que estas herramientas de IA permitan una protección cada vez más automatizada de los sistemas de tecnología de la información mediante una búsqueda de vulnerabilidades y actividades sospechosas que mejore la resiliencia de programas y equipos informáticos.

Al mismo tiempo, las herramientas de IA se utilizan crecientemente para realizar ciberataques más sofisticados y crear nuevos virus informáticos, en una carrera entre modelos ofensivos y defensivos de IA para la ciberseguridad. Además, los programas que funcionan con IA, como los grandes modelos lingüísticos, hacen más fácil el acceso de agentes malintencionados, que tienen más posibilidades de crear programas maliciosos sin necesidad de grandes conocimientos de programación.

Inteligencia artificial para campañas de desinformación en el marco de estrategias híbridas

Está aumentando el uso de programas informáticos basados en la IA con los que pueden crearse y difundirse contenidos tergiversados para amplificar campañas de desinformación. Métodos como la IA generativa permiten generar grandes volúmenes de contenidos localizados a medida. Además, el programa de ultrafalsificación de audio y vídeo con IA está progresando rápido y ya se utiliza ampliamente. Estos contenidos tergiversados pueden difundirse mediante redes masivas de bots en medios sociales que utilizan la IA para crear lo que parecen cambios en la opinión pública. Por tanto, la IA reduce los impedimentos para realizar grandes campañas de desinformación, ya que la cantidad y la calidad de los contenidos falsos creados ya no están limitadas por el número de operadores humanos o por sus competencias.

Sin embargo, los algoritmos de la IA también pueden servir para detectar contenidos generados con IA y estrategias de posverdad (*astroturfing*); resulta más fácil desenmascarar archivos ultrafalsificados de audio y el vídeo de apariencia real utilizando herramientas especializadas de IA. Es necesario emplear esas herramientas de IA para contrarrestar los efectos nocivos de la IA utilizada en campañas de desinformación.

Inteligencia artificial asociada a la proliferación de armas

La IA puede reducir las barreras de adquisición de armas, incluidas las de destrucción masiva. Al ofrecer conocimientos especializados con solo pulsar un botón, los grandes modelos lingüísticos y las aplicaciones que se basan en ellos podrían hacer que a los agentes malintencionados les resultara más fácil fabricar armas. Los usos son muy diversos, del acceso a prototipos o la impresión de componentes de armas pequeñas y armas ligeras a la modificación de patógenos para la guerra biológica. Si los conocimientos fácilmente disponibles reducen el alcance y el tamaño de los programas de armamento, será más difícil detectar, prevenir y prepararse para estas amenazas.

Al mismo tiempo, los algoritmos de aprendizaje automático también pueden utilizarse para luchar contra la proliferación de armas. Gracias a sus capacidades de detección de anomalías y de reconocimiento de patrones, pueden ayudar a descubrir actividades maliciosas, por ejemplo mediante la detección de flujos monetarios ilícitos destinados a programas de armamento o el análisis de patrones a partir de datos satelitales.

Inteligencia artificial para el control de armamentos y la toma de decisiones en situaciones de crisis

La IA puede ayudar a verificar acuerdos de control de armamentos gracias a su capacidad para analizar grandes cantidades de datos (de fuentes como las imágenes satelitales) y clasificar diferentes objetos, que permite detectar equipos militares, como tanques, misiles y cuarteles, o actividades militares, como desplazamientos de contingentes y maniobras. Además, como ya se ha mencionado, la IA permite detectar más fácilmente programas ilegales de armamento, lo que hará más difícil infringir los acuerdos de control de armas y permitirá asegurarse de que todos los Estados partes cumplen sus disposiciones.

Disponer de más y mejor información gracias a la capacidad de la IA de analizar y clasificar datos de sensores, además de facilitar la aplicación de los acuerdos de control de armamentos, puede también contribuir a mejorar la toma de decisiones en situaciones en las que las tensiones militares entre Estados son especialmente

elevadas. Los dirigentes políticos y militares podrían aprovechar la mayor conciencia situacional facilitada por la IA para la distensión de crisis.

La paz y la seguridad y la Carta de las Naciones Unidas

Entre los grandes retos que plantean las aplicaciones de la IA en el ámbito militar están los riesgos para la paz y la seguridad a los que pueden dar lugar la intensificación involuntaria de tensiones y los malentendidos generados al utilizar esta tecnología. El aprendizaje automático introduce un factor adicional de complejidad, ya que el funcionamiento de un sistema podría no ser plenamente comprendido por todos los actores.

También hacen falta medidas y salvaguardas que garanticen la rendición de cuentas y la responsabilidad y mitiguen los sesgos algorítmicos de uso de la IA en los sistemas de apoyo para la toma de decisiones en lo que respecta a la interacción humano-máquina y a la necesidad de intervención humana.

Todos estos riesgos deben mitigarse mediante actividades de supervisión y medidas que tengan en cuenta los retos específicos que entrañan estas tecnologías.

Cabe señalar que el artículo 36 del Protocolo Adicional a los Convenios de Ginebra del 12 de agosto de 1949 relativo a la Protección de las Víctimas de los Conflictos Armados Internacionales (Protocolo I) impone la obligación de examinar la legalidad de toda nueva arma, medio o método de guerra antes de utilizarlo en un conflicto armado.

La IA también puede utilizarse para reforzar el cumplimiento efectivo de las obligaciones del derecho internacional humanitario, en particular cuando se trata de la protección de civiles, como una obligación positiva y una acción afirmativa, incluso mediante proyectos, investigaciones y aplicaciones específicamente concebidos para ello.

Marcos de cooperación multilateral e intercambio de información

El desarrollo de la IA en el ámbito militar está siendo muy rápido y plantea retos a todos los Estados, por lo que los debates y foros multilaterales para intercambiar experiencias y mejores prácticas resultan de gran utilidad. Austria respaldó la Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar. Como Copresidenta del grupo de trabajo para el seguimiento de esta Declaración, Austria, junto con Alemania, ha facilitado el intercambio de mejores prácticas para abordar los retos y formular políticas en este ámbito. Austria también hizo suyo el Plan de Acción de la Cumbre sobre la Inteligencia Artificial Responsable en el Ámbito Militar y la Declaración de París sobre el Mantenimiento del Control Humano en los Sistemas de Armas que Funcionan con IA.

Relación entre la labor de la comunidad internacional en materia de inteligencia artificial en el ámbito militar y en materia de sistemas de armas autónomos

En lo que respecta globalmente a la aplicación de la IA y a la autonomía en el ámbito militar, cabe destacar la cuestión específica de los sistemas de armas autónomos, que plantean preocupaciones jurídicas, éticas y de seguridad específicas. Este no es el tema central de la resolución [79/239](#) de la Asamblea General, ya que en el marco de las Naciones Unidas lleva debatiéndose desde 2013 y un número de Estados cada vez mayor ha expresado el deseo de establecer normas y límites internacionales para los sistemas de armas autónomos. Por lo tanto, a los efectos del presente informe, Austria se limita a subrayar que prefiere un instrumento jurídicamente vinculante antes que sistemas de armas autónomos y se remite a la importante labor que está llevando a cabo el Grupo de Expertos Gubernamentales en

el marco de la Convención sobre Prohibiciones o Restricciones del Empleo de Ciertas Armas Convencionales que Puedan Considerarse Excesivamente Nocivas o de Efectos Indiscriminados, así como a las labores complementarias emprendidas a raíz de la primera resolución de la historia de la Asamblea General sobre los sistemas de armas autónomos letales (resolución 78/241), en virtud de la cual se publicó un informe del Secretario General (A/79/88), y de la resolución de seguimiento 79/62, en la que se decidió organizar consultas oficiales sobre los sistemas de armas autónomos letales, que se celebrarán en Nueva York los días 12 y 13 de mayo de 2025.

Consideraciones sobre los marcos jurídicos de inteligencia artificial

El Reglamento de Inteligencia Artificial es el marco jurídico establecido por la Unión Europea para los sistemas de IA en varios sectores y pretende fomentar la confianza en las aplicaciones de IA y el aprovechamiento de sus beneficios protegiendo al mismo tiempo los derechos humanos, las libertades fundamentales y los valores democráticos. Hace hincapié en la importancia de la transparencia, la responsabilidad y la supervisión humana en el desarrollo y el despliegue de los sistemas de IA, al tiempo que promueve la seguridad jurídica, la innovación y la competitividad. Aunque no se aplica a las aplicaciones de la IA destinadas a actividades militares, de defensa o de seguridad nacional, este Reglamento utiliza un enfoque basado en los riesgos, que podría resultar útil al abordar la gran diversidad de aplicaciones militares posibles de la IA.

Perspectivas para el futuro

Austria valora la labor desarrollada por las diferentes plataformas y foros mencionados en relación con las aplicaciones de la IA en el ámbito militar y confía en que contribuyan a elaborar un nuevo conjunto de normas y criterios concertados internacionalmente que garanticen un uso responsable de la IA en el ámbito militar que se atenga a las obligaciones jurídicas internacionales y a los principios éticos.

Chile

[Original: español]
[11 de abril de 2025]

Chile ha señalado anteriormente que el rápido desarrollo de tecnologías nuevas y emergentes constituyen un aspecto relevante en cuanto a la seguridad internacional, y representa un desafío para todos los países. Estas nuevas tecnologías, y en particular la inteligencia artificial (IA), pueden generar enormes beneficios para el desarrollo y bienestar de las sociedades, pero al mismo tiempo generar importantes interrogantes sobre las consecuencias de uso en el ámbito de la seguridad y la defensa. Las nuevas tecnologías pueden aportar importantes beneficios en cuanto a su uso, pero también riesgos y dificultades.

En tal sentido, Chile estima conveniente el poder generar un entendimiento común respecto al uso responsable de la IA en el ámbito militar y de la seguridad, como también respecto al desarrollo y uso de los denominados sistemas de armas autónomos letales. Chile apoya los esfuerzos multilaterales que generen y fortalecen las instancias de diálogo y discusión entre los países, objeto de poder lograr espacios de consenso y entendimiento mutuos sobre el uso de estas nuevas tecnologías.

Chile ha alcanzado una posición de liderazgo en el ámbito de la inteligencia artificial, tanto por sus importantes avances en las condiciones habilitantes para el despliegue de la tecnología como por sus progresos pioneros en materia de políticas y discusión regulatoria de la IA. En octubre de 2021, Chile lanzó su primera Política

Nacional de IA, desarrollada con la colaboración de diversos actores públicos y privados. Esta política se centra en tres ejes fundamentales, estos son: factores habilitantes, el uso y desarrollo de la tecnología, y el establecer marcos normativos y éticos que aseguren un uso responsable y seguro de la IA.

En el año 2024, Chile lanzó una versión actualizada de la Política Nacional de IA, incorporando nuevos sub-ejes como articulación internacional, medioambiente y crisis climática, inclusión y no discriminación, niños, niñas y adolescentes, y cultura y preservación del patrimonio cultural. La Política está acompañada de su respectivo plan de acción, que incluye más de 100 medidas comprometidas para el año 2026, abarcando áreas como educación, salud, medio ambiente y cultura. Asimismo, la nueva Política Nacional de IA incorporó los principios de la Recomendación sobre la Ética de la Inteligencia Artificial de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), con el fin de alinear el instrumento con los marcos internacionales más actualizados.

Cabe destacar que Chile fue el primer país del mundo en completar la metodología de evaluación del estadio de preparación, instrumento desarrollado por la UNESCO para medir el grado de preparación de un país para implementar la IA de manera ética y responsable, reafirmando su compromiso con la aplicación de la Recomendación sobre la Ética de la Inteligencia Artificial de la UNESCO en el ámbito normativo nacional. Chile ha impulsado el desarrollo ético y responsable de la tecnología, el cual se ha reflejado en la participación del país en las cumbres sobre la IA organizadas por el Reino Unido (2023), la República de Corea (2024) y Francia (2025).

En el ámbito legislativo, actualmente Chile discute un proyecto de ley para regular los sistemas de la IA, basado en el enfoque de riesgo, con el objetivo de promover su desarrollo e implementación de manera que respeten los principios democráticos y los derechos fundamentales de las personas.

En el ámbito de la defensa y la seguridad, Chile ha apoyado y participado activamente en las Cumbres sobre la Inteligencia Artificial Responsable en el Ámbito Militar, realizadas en La Haya (2023) y en Seúl (2024). Nuestro país ha respaldado los documentos aprobados en ambas cumbres (Call to action (2023) y Blueprint for action (2024)). De la misma forma, Chile apoya el trabajo de la Comisión global sobre la inteligencia artificial responsable en el ámbito militar.

Asimismo, cabe destacar que los días 13 y 14 de junio de 2024 se realizó en Chile un taller regional sobre el uso responsable de la inteligencia artificial en el ámbito militar y de seguridad en general, organizado por el Ministerio de Relaciones Exteriores de Chile y el Ministerio de Relaciones Exteriores de Costa Rica, y auspiciado por el Reino de los Países Bajos y la República de Corea. El evento contó además con el apoyo y organización del Centro de Estudios en Derecho, Tecnología y Sociedad de la Universidad de Chile, y del Centro para el Diálogo Humanitario, basado en Ginebra. Participaron en el taller representantes de Argentina, Brasil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Jamaica, México, Paraguay, República Dominicana, Trinidad y Tobago y Uruguay, así como el Reino de los Países Bajos y Corea. También por parte de Chile participaron representantes del Ministerio de Defensa Nacional y las Fuerzas Armadas.

Chile estima que la IA aplicada al ámbito militar y de seguridad puede generar oportunidades y beneficios, como, por ejemplo, mejorar la toma de decisiones y análisis estratégico, el llevar a cabo operaciones logísticas más eficientes, el mejorar las capacidades de ciberdefensa y de ciberseguridad fortaleciendo la seguridad de infraestructuras críticas, como también ayudar a planificar misiones complejas de mantenimiento de la paz y de ayuda humanitaria. También el mejorar las capacidades

de verificación y monitoreo en el control de armamentos y el cumplimiento de regímenes sobre el particular.

Chile estima que las tecnologías de IA deben desarrollarse, desplegarse y usarse de acuerdo con el derecho internacional, incluidos, según corresponda, la Carta de las Naciones Unidas, el derecho internacional humanitario, el derecho internacional de los derechos humanos y otros marcos legales relevantes.

Para Chile es importante que se puedan establecer medidas de control y de seguridad que eviten que actores irresponsables adquieran y hagan uso indebido de las capacidades de la IA potencialmente perjudiciales en el ámbito militar, incluidos los sistemas habilitados por la IA, teniendo en cuenta que estas medidas no deben socavar el acceso equitativo a los beneficios de las capacidades de la IA en otros dominios no militares.

De la misma forma, Chile estima que es importante aunar esfuerzos para prevenir que las tecnologías de la IA se utilicen para contribuir a la proliferación de armas de destrucción masiva por parte de actores estatales y no estatales, incluidos grupos terroristas, y enfatizar que las tecnologías de la IA deben apoyar y no obstaculizar los esfuerzos de desarme, control de armamentos y no proliferación. Es especialmente crucial mantener el control humano y la participación en todas las acciones críticas para informar y ejecutar decisiones soberanas relacionadas con el empleo de armas nucleares, sin perjuicio del objetivo final de un mundo libre de armas nucleares.

Chile alienta el desarrollo de medidas de fomento de la confianza, como el intercambio de información y consultas sobre buenas prácticas y lecciones aprendidas entre los Estados. En tal sentido, para Chile es importante que los países puedan desarrollar y establecer estrategias nacionales, principios, normas, políticas, marcos y legislación para garantizar aplicaciones responsables de la IA en el ámbito militar. Las medidas de fomento de la confianza pueden ser una herramienta eficaz para desarrollar mecanismos de contención, control y credibilidad, tanto a nivel nacional como internacional, fomentando la transparencia.

De la misma forma, Chile considera que es fundamental reducir las brechas digitales y de la IA que existen entre los países desarrollados y los países en desarrollo, y estima que es necesario mejorar la comprensión y la sensibilización respecto de las implicaciones que tiene la inteligencia artificial en el ámbito militar, incluyendo el intercambio de conocimientos y la puesta en común de buenas prácticas y lecciones aprendidas entre todos los Estados.

En ese sentido, Chile estima que es esencial el poder desarrollar iniciativas y programas que fomenten el desarrollo de capacidades, especialmente en los países en desarrollo, para promover la plena participación de dichos países en los debates sobre la gobernanza de la IA en el ámbito militar, reconociendo que el desarrollo de capacidades también puede ayudar a los países a profundizar su comprensión de la IA en el ámbito militar y a facilitar el desarrollo, el despliegue y el uso de las capacidades militares de la IA de forma responsable y legal. La creación de capacidades permitirá también a los países el poder enfrentar de mejor manera la discusión y el diálogo internacionales.

Chile considera importante fortalecer la cooperación internacional en materia de desarrollo de capacidades, promoviendo el diálogo y debate a nivel nacional, regional, subregional e interregional, incluyendo programas de capacitación, conferencias, talleres y seminarios, entre otros, para funcionarios a nivel diplomático, político y técnico, con el fin de reducir la brecha de conocimiento sobre el desarrollo, el despliegue y el uso responsables de la IA en el ámbito militar.

Chile valora y estima importante promover la discusión y el diálogo regional y subregional sobre la IA en el ámbito militar, y en ese sentido destaca la XVI Conferencia de Ministros de Defensa de las Américas, la cual se realizó en la Argentina del 13 al 16 de octubre de 2024, y en particular la Declaración de Mendoza, documento emanado de dicha Conferencia, y en el cual se recomienda, entre otros, el fomentar el uso ético de la IA en defensa; el considerar la diversidad económica y tecnológica de los países miembros de la Conferencia; y el promover mecanismos de fortalecimiento de la confianza mutua y de cooperación hemisférica y regional que permitan a los países miembros de la Conferencia compartir conocimientos y buenas prácticas y construir estándares consensuados, así como desarrollar capacidades tecnológicas para la aplicación de la IA en el ámbito de la defensa.

Finalmente, Chile estima que, en la discusión y diálogo sobre la IA en el ámbito militar, es esencial contar con la participación de todas las partes interesadas, como por ejemplo la sociedad civil, la academia, la industria, el sector privado, la comunidad técnica y organizaciones regionales e internacionales, entre otros.

China

[Original: chino]
[11 de abril de 2025]

La rápida evolución y el amplio uso de la inteligencia artificial (IA) en el ámbito militar están redefiniendo los paradigmas bélicos del futuro, aunque también pueden conllevar problemas para la paz y la seguridad internacionales. Habida cuenta de que el mundo afronta múltiples retos para la paz y la seguridad, todas las partes deben tratar de consensuar, mediante el diálogo y la cooperación, unas normas que regulen las aplicaciones militares de la IA, promover el desarrollo de un marco de gobernanza abierto, justo y eficaz que dé seguridad a la IA y reducir al mínimo los riesgos para que estas tecnologías sigan siendo seguras, fiables y controlables y se desarrollen siempre de forma que contribuyan al progreso de la civilización humana.

China siempre ha participado de forma responsable y constructiva en la gobernanza global de las aplicaciones militares de la IA. En nuestra apuesta por forjar una comunidad humana con un futuro compartido, propugnamos que se asuma un “enfoque centrado en las personas en las aplicaciones militares de la IA” y se defienda la idea de una seguridad integral, cooperativa y sostenible común. En 2021 China presentó, en el marco de la Convención sobre Ciertas Armas Convencionales, un documento de posición sobre la regulación de las aplicaciones militares de la IA en el que proponía perspectivas y recomendaciones sistemáticas sobre el desarrollo y el uso responsables de la IA en el ámbito militar en lo que se refiere a la seguridad estratégica, las políticas militares, el derecho y la ética, la seguridad tecnológica, las actividades de investigación y desarrollo, la gestión y el control de riesgos, el establecimiento de normas y la cooperación internacional. En 2023, China presentó la Iniciativa Mundial de Gobernanza de la IA, en la que se instaba a todos los países, y en especial a las grandes potencias, a adoptar una actitud prudente y responsable respecto de la investigación, el desarrollo y la aplicación de las tecnologías de IA en el ámbito militar. A continuación se indican algunas de nuestras propuestas concretas:

En primer lugar, debería adoptarse un enfoque prudente y responsable. Al desarrollar sus capacidades legítimas de defensa nacional, todos los países, y en particular las grandes potencias, deberían evitar utilizar la IA para conseguir una superioridad militar absoluta y socavar los intereses de seguridad legítimos de los demás. Habría que poner empeño en evitar malentendidos y errores de cálculo y prevenir una carrera armamentista en este ámbito.

En segundo lugar, debería aplicarse un enfoque centrado en las personas. Es indispensable considerar en todo momento al ser humano como responsable último y asegurarse de que los sistemas de armamento en cuestión estén bajo su control. Las aplicaciones militares de la IA deben respetar y proteger la dignidad y los derechos humanos y honrar los valores comunes de la humanidad.

En tercer lugar, debería observarse el principio básico de “la IA para el bien”. Las aplicaciones militares de la IA han de contribuir a mantener la paz y cumplir el derecho internacional humanitario y otras normas internacionales vigentes y procurar reducir las bajas colaterales.

En cuarto lugar, debería instaurarse una gobernanza ágil. Hemos de reforzar la evaluación prospectiva de los riesgos y la capacitación del personal en materia de IA, adoptar las medidas de mitigación del riesgo que hagan falta y reducir las posibilidades de proliferación, sin entorpecer con ello la innovación y los usos pacíficos de las tecnologías.

En quinto lugar, hay que defender el multilateralismo. Tendríamos que ayudar a las Naciones Unidas a desempeñar la función que le corresponde, celebrar la creación de plataformas inclusivas que permitan debatir a todas las partes y tratar de establecer marcos de gobernanza que se elaboren con participación de todos y gocen de amplio consenso.

China considera necesario evaluar de manera objetiva la importancia de la IA en el ámbito militar. Es imprescindible orientar en un sentido adecuado la evolución de la IA y evitar que se desarrolle sin orden ni concierto. En la siguiente etapa, la comunidad internacional debería colaborar para sacar el máximo provecho de esta tecnología y reducir al mínimo los daños. A continuación se presentan algunas ideas y propuestas de China:

Primero, establecer directrices claras. Ha de atenderse por igual a la seguridad y al desarrollo. Es indispensable respetar los propósitos y principios de la Carta de las Naciones Unidas, observar las normas básicas que rigen las relaciones internacionales y velar por que la tecnología de la IA no se convierta en una herramienta para invadir a otros países y buscar la hegemonía. China está deseosa de seguir abordando con todas las partes el concepto de “enfoque centrado en las personas en las aplicaciones militares de la IA” y de fomentar constantemente el consenso.

Segundo, mejorar las medidas de gobernanza. Habida cuenta del actual estado de desarrollo y aplicación de la IA, deberíamos promover el establecimiento de un sistema de pruebas y evaluación, instaurar una gobernanza ágil e implantar una gestión por niveles y categorías que permita responder con rapidez y eficacia. En función de sus condiciones, todos los países deberían establecer sistemas jurídicos y reguladores nacionales y perfeccionarlos, afinar las directrices éticas en la materia y reforzar la educación y la capacitación a fin de mejorar la seguridad, la fiabilidad y la controlabilidad de las tecnologías de IA.

Tercero, fortalecer la cooperación internacional. Todos los países deben adherirse a los principios de apertura e inclusión, dialogar e intercambiar pareceres a fin de promover el entendimiento mutuo y poner mayor empeño en coordinar las políticas y en cooperar en la creación de capacidades de gobernanza de la IA para hacer que mejore continuamente.

Egipto

[Original: inglés]
[11 de abril de 2025]

De conformidad con la resolución [79/239](#) de la Asamblea General, el Gobierno de la República Árabe de Egipto desea dar a conocer sus opiniones sobre las oportunidades y retos para la paz y la seguridad internacionales que plantea la aplicación de la inteligencia artificial (IA) en el ámbito militar.

La resolución [79/239](#) de la Asamblea General da un paso importante para fomentar el multilateralismo en torno a la aplicación de la IA en el ámbito militar y para situar esta cuestión en un lugar más destacado de la agenda política. Responde al llamamiento del Secretario General a elaborar normas, reglas y principios de concepción, desarrollo y utilización de las aplicaciones militares de la IA con la participación de todas las partes interesadas pertinentes.

En el entendimiento de que dicha resolución, en virtud de la cual se presentan estas opiniones, pretende centrarse específicamente en ámbitos distintos de los sistemas de armas autónomos letales, cabe reiterar la firme posición de Egipto de que cualquier debate sustancial sobre este tema no puede desatender la prioridad de abordar todas las cuestiones éticas, jurídicas y de seguridad que plantean los sistemas de armas autónomos letales, que constituyen la amenaza más acuciante para el mantenimiento de la paz y la seguridad internacionales que se deriva de las aplicaciones militares de la IA.

Acordar una prohibición jurídicamente vinculante de los sistemas de armas autónomos letales que funcionen sin control o supervisión humana y que no puedan utilizarse de conformidad con el derecho internacional humanitario, como ya planteó el Secretario General, es la línea de acción más eficaz y realista. Es esencial aplicar un planteamiento doble, de prohibición y restricción y de regulación (prohibición de los sistemas de armas que funcionen sin control humano y regulación de los demás), a fin de crear la arquitectura jurídica universal necesaria para generar unas condiciones que permitan aprovechar al máximo las ventajas y nuevas oportunidades que ofrecen las aplicaciones militares de la IA, al tiempo que se abordan los retos conexos de forma realista, eficaz y oportuna.

En lo que respecta a la aplicación militar de la IA, las diversas políticas internacionales distan mucho de estar unificadas. Egipto sigue de cerca las múltiples iniciativas internacionales sobre el particular, que ponen de relieve un grado de conciencia cada vez mayor de los riesgos asociados a esta tecnología. Sin embargo, las deliberaciones mantenidas a raíz de esas iniciativas han revelado divergencias en los puntos de vista, las percepciones de las amenazas y las prioridades, por lo que cabe alertar del peligro de crear un marco político fragmentario o procesos concurrentes, como ocurrió con otras tecnologías nuevas y emergentes.

Existe una clara necesidad de racionalizar esas iniciativas y situarlas bajo el marco de las Naciones Unidas para garantizar su inclusividad y eficacia. Las Naciones Unidas y su maquinaria de desarme son la única plataforma eficaz e inclusiva para elaborar las reglas y el marco normativo internacionales que se necesitan, sobre todo teniendo en cuenta que el desarrollo tecnológico sigue siendo mucho más rápido que el establecimiento de la regulación internacional que hace falta.

De ahí la necesidad imperiosa de crear, bajo los auspicios de las Naciones Unidas, un foro universal, independiente, unificado y digno de crédito para abordar la futura gobernanza de la IA en el ámbito militar. El proceso previsto bajo la dirección de las Naciones Unidas deberá ajustarse para despejar ciertas dicotomías contraproducentes que empiezan a plantearse, como la que opone las labores

legítimas por garantizar el cumplimiento de las normas jurídicas y éticas a la tendencia a primar intereses militares sin tener debidamente en cuenta las implicaciones humanitarias.

Cabe destacar asimismo que, aun valorando los debates sobre los sistemas de armas autónomos letales que el Grupo de Expertos Gubernamentales mantiene en el marco de la Convención sobre Prohibiciones o Restricciones del Empleo de Ciertas Armas Convencionales que Puedan Considerarse Excesivamente Nocivas o de Efectos Indiscriminados, esa plataforma no puede remplazar el proceso previsto en las Naciones Unidas sobre las aplicaciones de la IA en el ámbito militar, ya que dicho Grupo no es de carácter universal y carece de mandato para tratar un tema tan multidimensional y diverso. También es de lamentar que los avances del Grupo de Expertos Gubernamentales sigan siendo más bien escasos y que aún no haya dado ningún resultado concreto.

Las tecnologías de IA ofrecen oportunidades, pero sus características conllevan una serie de riesgos que provocar un funcionamiento impredecible e inexplicable. Entre esos riesgos están la desinformación, la intensificación indeseada de tensiones y los ciberriesgos, así como el uso indebido y la proliferación por parte de actores no estatales. Puede tratarse de riesgos nuevos o de factores que vienen a complicar riesgos ya existentes.

Aunque, según se reconoce ampliamente, las posibles aplicaciones militares de la IA son muy diversas, para trabajar de manera efectiva en la concepción de su gobernanza futura habrá que establecer un orden de prioridades adecuado en función de los riesgos que entraña cada una de ellas y de sus consecuencias para la paz y la seguridad, lo que permitirá centrar y estructurar los debates evitando distracciones indebidas. Dicho esto, Egipto tiene la firme opinión de que el énfasis debe ponerse, aparte de la cuestión de los sistemas de armas autónomos letales, en otras capacidades de sistemas autónomos o semiautónomos que permitan hacer uso de la fuerza o rebajen los criterios para hacerlo, lo que podría generar una nueva carrera armamentista que afecte tanto a las armas convencionales como las no convencionales. La posibilidad de una mayor autonomía de las armas nucleares y de armas convencionales avanzadas, como los misiles hipersónicos, entrañaría riesgos desconocidos y alteraría de forma impredecible el futuro de los conflictos.

También habrá que prestar a las actividades de mando y control y de selección de objetivos más atención que a las de planificación logística e inteligencia, vigilancia y reconocimiento, cuyos efectos son menos perturbadores. De manera análoga, habrá que centrarse más en las capacidades ofensivas que en las defensivas.

Las deliberaciones previstas en el marco del deseado proceso dirigido por las Naciones Unidas tendrán como primer objetivo alcanzar un entendimiento común sobre los principales elementos en que se asientan el desarrollo, el despliegue y el uso de la IA en el ámbito militar, en particular los siguientes:

- El pleno respeto del derecho internacional vigente, incluidos principios cardinales del derecho internacional humanitario como los de necesidad, proporcionalidad y distinción, así como de otras consideraciones éticas en todas las etapas del ciclo de vida de las aplicaciones militares de la IA.
- El mantenimiento de la centralidad del factor humano en todo el ciclo de vida de las aplicaciones militares de la IA, que comprende el juicio, la intervención, la supervisión y el control humanos, que son las principales garantías de que siga habiendo responsabilidad. Es necesario garantizar que todos los programas informáticos, algoritmos y diseños que impliquen la utilización de tecnologías de IA en el ámbito militar sigan estando sujetos a examen por parte de seres humanos y al principio de explicabilidad. Aunque los Gobiernos sostienen que

la prevalencia del control humano sobre los sistemas dotados de IA es un principio inmutable, algunos pueden caer en la tentación de hacer que sus sistemas de armas sean cada vez más autónomos para favorecer sus intereses militares.

- El equilibrio entre la reducción de los riesgos de proliferación entre agentes no estatales y de utilización malintencionada, por una parte, y la preservación del derecho de los Estados a adquirir tecnologías de IA y doble uso, por otra. Es crucial evitar la introducción de mecanismos de supervisión internacional arbitrarios o la imposición de cualquier tipo de control discriminatorio de las exportaciones.
- Un componente de creación de capacidad que permita una inversión adecuada en capital humano, la transferencia de tecnología y el intercambio de conocimientos y mejores prácticas a fin de proteger el derecho de los países en desarrollo a aprovechar las ventajas potenciales de las diversas aplicaciones militares de la IA y reducir la brecha digital.
- Los límites de la IA en el ámbito militar y su interacción con otras tecnologías nuevas y emergentes. Conviene buscar vías de complementariedad con otros procesos promovidos por las Naciones Unidas, como el grupo de trabajo de composición abierta sobre la seguridad de las tecnologías de la información y las comunicaciones y de su uso, habida cuenta de las intersecciones entre la IA y las ciberoperaciones. Además, los debates se centrarán principalmente en el ámbito militar, dejando al margen otros aspectos de la seguridad en general.

Por último, al establecer medios de gobernanza para una IA responsable, que entraña rendición de cuentas y se centre en el ser humano, es importante acoger de manera inclusiva y equitativa las perspectivas de las múltiples partes interesadas de las Naciones Unidas que hacen aportaciones esenciales para informar los debates sobre políticas. No obstante, esa participación no afectará a la prerrogativa soberana de los Estados en el proceso de elaboración de políticas.

El Salvador

[Original: español]
[10 de abril de 2025]

Antecedentes

Durante los últimos años, el uso de la inteligencia artificial (IA) en el dominio militar ha tenido un papel muy importante. Numerosos informes establecen que estas nuevas tecnologías son cada vez más sofisticadas y generalizadas, permitiendo que estas herramientas informáticas sean utilizadas en procesos de planificación y toma de decisiones militares, incluso sobre quién o qué atacar. Lo anterior genera múltiples interrogantes en cuanto a las consecuencias generales, las implicaciones legales y los riesgos hacia la población civil del uso de estas tecnologías. Ejemplo de ello, ha sido el debate acerca de las implicaciones de estas tecnologías en la negociación multilateral centrado en los sistemas de armas autónomos, desde perspectivas políticas, jurídicas y humanitarias. Sin embargo, se reconoce que las aplicaciones militares de la IA son mucho más amplias.

Hay por lo anterior una necesidad de aumentar nuestra comprensión del uso y aplicaciones de la IA que se utilizan en el contexto militar, particularmente para las tareas específicas de la selección de objetivos militares y la aplicación de la fuerza.

El tema del uso responsable de la IA en el dominio militar ha cobrado especial relevancia a partir de las discusiones desarrolladas en la primera Cumbre sobre la

Inteligencia Artificial Responsable en el Ámbito Militar que tuvo lugar en el Reino de los Países Bajos en febrero de 2023. Adicionalmente, este tema también comenzó a cobrar relevancia en las sesiones del Grupo de Expertos Gubernamentales sobre los Sistemas de Armas Autónomos Letales que sesiona en Ginebra.

Si bien es relevante mencionar que las aplicaciones y usos de la IA han sido debatidos principalmente en las discusiones sobre los sistemas de armas autónomos hasta este momento, el uso de la IA para aplicaciones militares es mucho más amplio y está cobrando una nueva dimensión, particularmente para aquellas aplicaciones orientadas a la automatización de algunas funciones militares y no necesariamente solo aquellas orientadas a la autonomía en los sistemas de armas.

En términos generales, la discusión sobre la IA es un tema emergente del cual todavía se sigue investigando y que está avanzando muy rápidamente. Tanto así que se están generando iniciativas a nivel nacional, regional y multilateral para poder abordar su impacto. Es claro que los países de América Latina y el Caribe no tienen el mismo nivel de tecnología y desarrollo de capacidades que permita determinar y comprender las oportunidades y riesgos que derivan del uso de la IA, como lo tienen los países desarrollados. Por ello, es importante poder definir una posición nacional que les permita participar activamente en las discusiones que están surgiendo en los espacios y foros internacionales y lograr con ello cooperación en generación de capacidades y aspectos específicos que permitan estar la vanguardia en el tema y comprender las oportunidades y posibles riesgos en materia de seguridad a nivel nacional, regional y global.

Iniciativas de las que El Salvador ha participado

- El Salvador participó en la Cumbre sobre la Inteligencia Artificial Responsable en el Ámbito Militar, que tuvo lugar en 2023 en el Reino de los Países Bajos y se sumó a la Declaración que de ella emanó (febrero de 2023).
- Participó en la Conferencia Latinoamericana y del Caribe sobre el Impacto Social y Humanitario de las Armas Autónomas, donde se adoptó el Comunicado de Belén (febrero de 2023).
- El Salvador es parte del “Grupo de 16” en el marco de las discusiones del Grupo de Expertos sobre los Sistemas de Armas Autónomos Letales. Si bien este es un tema con su propia visión, tiene relación con el uso de la IA en el ámbito militar.

Posición nacional

- Algunas aplicaciones de la IA pueden tener ciertos beneficios en el campo militar, particularmente aquellas que no se asocian con las funciones de identificación y reconocimiento de blancos militares y precisamente para el uso de la fuerza, al presentar un riesgo para la población civil, sino aquellas relacionadas a otras labores administrativas, como análisis de datos y aprendizaje automático no relacionado con interacción humana en operaciones militares.
- Sin embargo, un mal uso de estas aplicaciones puede tener efectos adversos, particularmente relacionados a la protección de civiles e infraestructura civil, que tiene categorías de protección especial basadas en las normas del derecho internacional, entre ellas el derecho internacional humanitario y el derecho internacional de los derechos humanos.
- Es importante contar con un enfoque basado en riesgos que permita la regulación y prohibición de algunas funciones de la IA, particularmente aquellas que restringen el control humano significativo en el uso de la fuerza, aquellas

que perpetúan sesgos algorítmicos debido al uso de bases de datos poco representativas o con datos históricos y representan riesgos relacionados con los derechos humanos y, a largo plazo, representan riesgos para la seguridad internacional, en especial cuando se le brinda la potestad a una máquina para decidir entre la vida y la muerte de un ser humano, o cuando estas herramientas poseen elementos tecnológicos muy sofisticados, como el autoaprendizaje, que podrían tener graves implicaciones humanitarias, sociales, económicas, políticas e incluso medioambientales.

- Actualmente urge la incorporación de una regulación adecuada en el campo de la IA, pues es esencial para garantizar su desarrollo ético y seguro, que permitirá la protección a los usuarios y a la sociedad de posibles abusos y riesgos, así como también fomentará la innovación al proporcionar un entorno claro y seguro para los desarrolladores e investigadores.
- Si bien el objetivo final es la formulación de instrumentos jurídicamente vinculantes, se considera que el avance de estas tecnologías se desarrolla de una manera más rápida con relación a la evolución o elaboración del derecho internacional en este ámbito, por lo que consideramos oportuno mantener un enfoque que se centre en el comportamiento responsable, sobre el cual luego se permita construir la base de compromisos jurídicos integrales para un mejor abordaje de este tema.
- Es importante considerar el desafío que imponen las tecnologías emergentes en las cuestiones de seguridad; tal es el caso del uso de tecnologías de materiales, como impresión 3D, para la fabricación de armas pequeñas y ligeras, el uso de la robótica para el desarrollo de robots con capacidades autónomas en el campo militar, así como algunos usos y aplicaciones de la IA, que por su naturaleza dual, pueden replicar sesgos en funciones de comando y control en conflicto armado, particularmente representando riesgos incrementados para civiles.
- La pérdida o sustitución de control del dominio militar puede generar riesgos no intencionales. La IA puede incrementar capacidades humanas, pero la falta de control en el contexto militar puede tener otros riesgos, que deben explorarse a profundidad. La asistencia de la IA en el dominio militar debería reforzar o informar la toma de decisiones en contextos específicos, pero jamás reemplazar la decisión y raciocinio humano.
- El uso de IA en el ámbito militar deberá cumplir con el derecho internacional, el derecho internacional de los derechos humanos y el derecho internacional humanitario y servir al bien público.
- Es necesario que se puedan fortalecer las capacidades de los países para poder identificar los riesgos derivados de un mal uso de la IA y su vinculación con el derecho internacional.
- Se deben incluir en las discusiones multilaterales a otros actores que son parte del proceso de creación y desarrollo de este tipo de tecnología, como por ejemplo las empresas y la academia, entre otros, así como fomentar la cooperación internacional entre las partes involucradas, para aprovechar los beneficios que ofrece la IA para fines pacíficos a favor del desarrollo de los países.

España

[Original: español]
[11 de abril de 2025]

Introducción

La inteligencia artificial (IA) supone una revolución en todos los ámbitos, y también lo es para la seguridad y la defensa. Su desarrollo e integración conlleva consigo grandes avances, oportunitades, a la vez que plantea numerosos retos.

La adopción de esta tecnología en las fuerzas armadas no solo redefine la forma en que se llevan a cabo las operaciones militares, sino que también transforma el equilibrio estratégico global.

El desarrollo e implantación de la IA en el Ministerio de Defensa se basa en un uso militar responsable, ético y legítimo acorde al derecho internacional humanitario y al respeto a los derechos humanos.

La IA está modificando la concepción tradicional del poder militar y la seguridad, proporcionando capacidades avanzadas para la recopilación y análisis de datos, la toma de decisiones y la ejecución de operaciones en entornos multidominio. Esto implica un cambio de paradigma en la forma en que los Estados abordan la defensa y la seguridad, facilitando respuestas más rápidas y precisas ante amenazas emergentes.

En el ámbito militar, la IA está impactando de manera disruptiva en el imprevisible campo de batalla, lo que significa un cambio de paradigma en el planeamiento y conducción de las operaciones militares. También está impactando en otras cuestiones dentro del ámbito militar: logística, entrenamiento, gestión e interpretación de información, inteligencia, vigilancia, adquisición de objetivos y reconocimiento.

Coherentemente con el compromiso de España con el uso responsable de la IA, cabe destacar la organización de la Cumbre sobre Inteligencia Artificial Responsable en el Ámbito Militar de 2025 como nación anfitriona, además de su adhesión a la “Llamada a la Acción” (La Haya, 2023) y el “Plan de Acción” presentado en la pasada Cumbre en 2024.

Marco conceptual y normativo en el Ministerio de Defensa

El desarrollo, despliegue y aplicación de la IA en el Ministerio de Defensa están guiados por un conjunto de principios fundamentales que garantizan su uso seguro, ético y conforme a las normativas nacionales e internacionales. Estos principios, que se encuentran compilados en la “Estrategia de desarrollo, implantación y uso de la Inteligencia Artificial en el Ministerio de Defensa” (desarrollada mediante la resolución 11197/2023 de la Secretaría de Estado de Defensa) y están alineados con la estrategia de la Organización del Tratado del Atlántico Norte (OTAN) para la IA (2021, revisada en 2024), buscan maximizar las oportunidades que ofrece la IA para la defensa, al mismo tiempo que buscan mitigar los riesgos asociados a su implementación en el ámbito militar, desglosados a continuación.

- Legalidad: las aplicaciones de la IA se desarrollarán y emplearán de acuerdo con el derecho nacional e internacional que sea de aplicación, incluida la Declaración Universal de Derechos Humanos y el derecho internacional humanitario.
- Responsabilidad humana y rendición de cuentas: cualquier desarrollo y utilización de la IA deberá permitir una clara supervisión humana con el fin de garantizar la debida rendición de cuentas y la atribución de responsabilidades.

- Inteligibilidad y trazabilidad: las aplicaciones de la IA serán comprensibles y transparentes para el personal relevante, incluyendo el uso de metodologías, fuentes y procedimientos auditables.
- Fiabilidad y transparencia: las aplicaciones de la IA estarán dirigidas a casos de uso explícitos, bien definidos y acotados y se proporcionará información para fomentar una comprensión general de estas aplicaciones por todas las partes interesadas. La seguridad, la protección y la solidez de estas capacidades estarán sujetas a pruebas y garantías dentro de estos casos de usos a lo largo de todo su ciclo de vida.
- Gobernabilidad: las aplicaciones de la IA se desarrollarán y utilizarán de acuerdo con sus funciones previstas en el diseño, asegurando la capacidad de detectar y evitar consecuencias no deseadas. Se habilitarán mecanismos de desconexión o desactivación cuando se reconozcan comportamientos no previstos o indeseados.
- Mitigación del sesgo: se tomarán todas las medidas necesarias para minimizar errores y orientaciones subjetivas en el desarrollo y aplicaciones de la IA.
- Privacidad: el desarrollo, la implementación y el uso de aplicaciones basadas en la IA deben respetar la privacidad de las personas, desde el diseño y durante todo el ciclo de vida.

Con respecto al marco normativo, para el desarrollo, implementación y uso de la IA en el ámbito militar dentro del Ministerio de Defensa se están elaborando un conjunto de normas y buenas prácticas que garantizan su aplicación responsable, eficiente y alineada con los marcos legales nacionales e internacionales, y en particular con el estricto cumplimiento con el derecho internacional humanitario y los derechos humanos.

Oportunidades

El Ministerio de Defensa enfoca el desarrollo de sus capacidades en la IA en múltiples áreas para mejorar la eficacia de las fuerzas armadas. Según la estrategia, el uso de la IA está enfocado a los ámbitos de operaciones, inteligencia, logística y ciberseguridad, así como al apoyo a la toma de decisiones.

La IA permitirá incrementar la precisión, rapidez y efectividad en la toma de decisiones durante una operación militar siempre sujeta al derecho internacional humanitario, en pos de una ejecución más eficiente de las misiones y reduciendo el riesgo para las tropas, así como contribuir a aumentar la protección de civiles y de los bienes de carácter civil en los conflictos armados.

Su capacidad para analizar grandes volúmenes de datos en tiempo real mejora la conciencia situacional y la capacidad de respuesta frente a amenazas, lo que refuerza la seguridad operativa. Toda esta mejora de capacidades siempre tiene presente el control humano, no delegando la responsabilidad en las máquinas.

En cuanto a formación y enseñanza militar, en el marco de la colaboración del grupo de Escuelas de Estado Mayor europeas (“C5 Commandants Group”) (Gran Bretaña, Francia, Alemania, Italia y España), se está trabajando en la creación de un área de colaboración relativa a la IA en la enseñanza militar.

A su vez, se colabora con la Junta de Revisión de Datos e Inteligencia Artificial de la OTAN para el uso responsable de datos y de la IA en el ámbito militar.

Por otra parte, el Ministerio de Defensa ha anunciado inversiones estratégicas en regiones específicas con el objetivo de impulsar proyectos relacionados con la IA

y otras tecnologías avanzadas. Estas inversiones buscan no solo fortalecer la industria, sino también promover la revitalización industrial de nuevas zonas regionales.

Retos

El desarrollo y aplicación de la IA en el ámbito militar deben alinearse con marcos normativos nacionales e internacionales, incluyendo el respeto al cumplimiento del derecho internacional humanitario, de manera que se está trabajando para asegurar un control humano efectivo en decisiones críticas asociadas al empleo de la IA en operaciones militares.

Por otra parte, desde una perspectiva de privacidad y protección de datos, la recopilación y procesamiento masivo de datos para entrenar modelos de IA plantea riesgos en materia de protección de datos personales y seguridad de la información.

Seguridad y fiabilidad

El mayor reto se plantea en relación con el empleo seguro y fiable de la IA, siendo los principales riesgos asociados los siguientes:

- Los algoritmos de la IA pueden presentar sesgos en los datos de entrenamiento, lo que puede llevar a decisiones erróneas o con consecuencias no deseadas.
- Un mal entrenamiento de los modelos de la IA puede generar errores de interpretación, con consecuencias potencialmente catastróficas en operaciones militares.
- Los sistemas de la IA pueden ser objetivo de ataques ciberneticos, que pueden manipular su comportamiento o inutilizarlos.
- Existe el riesgo de envenenamiento de datos, donde actores malintencionados alteran los conjuntos de datos de entrenamiento para provocar fallos en los algoritmos.

En España, el desarrollo de la IA en el ámbito militar se rige por los principios de responsabilidad y supervisión continua, implementando mecanismos de evaluación de riesgos, auditoría y trazabilidad en cada fase del ciclo de vida del sistema. Cualquier desarrollo de la IA, así como su utilización, deberá permitir una clara supervisión humana con el fin de garantizar la debida rendición de cuentas y la atribución de responsabilidades, dejando clara trazabilidad de las actuaciones humanas relacionadas y tomadas en paralelo con la actuación de la IA, y no delegando la decisión final en máquinas.

La IA debe ser además fiable y previsible, manteniendo un nivel de autonomía controlado y supervisado por operadores capacitados.

Cualquier solución de la IA se revisará en un entorno diferente al que fue entrenado, y será sometido a pruebas no funcionales: carga, estrés y rendimiento bajo escenarios cambiantes definidos, para estudiar su comportamiento y desviación permitida.

Además, estas capacidades de la IA estarán sujetas a pruebas rigurosas y auditorías constantes a lo largo de todo su ciclo de vida, permitiendo la detección temprana de posibles errores y mejorando su fiabilidad operativa. Se implementarán protocolos de supervisión y control humano en todas las fases de despliegue, asegurando que las decisiones críticas no sean delegadas exclusivamente a la IA. En este sentido, se está trabajando para que los desarrollos de la IA sean certificados por entidades reconocidas.

Con objeto de mejorar la robustez de los sistemas basados en la IA y protegerlos frente a acciones externas, es clave integrar la seguridad desde el diseño, garantizando que estos sistemas sean resistentes a ciberataques y manipulaciones adversarias, para asegurar la integridad de los datos y modelos utilizados.

La IA puede ser objetivo de ataques como el envenenamiento de datos o la manipulación de modelos, por lo que se requiere una monitorización continua del desempeño de los sistemas y la implementación de pruebas de validación y auditorías regulares. Se fomentará el desarrollo de planes de respaldo y recuperación ante fallos, garantizando la operatividad de los sistemas en escenarios adversos.

Se promueve también la colaboración con organismos de ciberseguridad y expertos en la IA, asegurando que las fuerzas armadas cuenten con las mejores herramientas y estrategias para proteger estos sistemas de amenazas externas y asegurar su fiabilidad operativa.

Por otra parte, el talento y la formación del personal en estas tecnologías es crítico por lo que es contemplado como uno de los cuatro pilares en los que pone el foco el Ministerio de Defensa, asegurando que los operadores comprendan el alcance y las limitaciones de estos sistemas y puedan intervenir en caso de desviaciones en su comportamiento. La formación y sensibilización del personal en el uso legal y ético de la IA es clave para mitigar los riesgos asociados a los sesgos, asegurando que su empleo en las fuerzas armadas sea objetivo, fiable, y conforme a las normativas nacionales e internacionales, especialmente con el derecho internacional humanitario.

Actualmente se está trabajando en una guía de buenas prácticas que podría ser la base de un documento al que contribuyan todos los ámbitos del Ministerio de Defensa. Se han divulgado aquellas buenas prácticas que desde la OTAN se proponen para el uso responsable de la IA en el ámbito militar. Un ejemplo es el conjunto de herramientas de la OTAN para la evaluación responsable de la IA (NATO Responsible AI (RAI) Assessment and Toolkit), cuyo propósito es operacionalizar los principios de uso responsable de la IA adoptados por la OTAN, que incluyen legalidad, responsabilidad, trazabilidad, fiabilidad, gobernabilidad y mitigación de sesgos.

Federación de Rusia

[Original: ruso]
[10 de abril de 2025]

La Federación de Rusia acoge con satisfacción la aprobación de la resolución 79/239 de la Asamblea General de las Naciones Unidas, de 24 de diciembre de 2024, y, de conformidad con su párrafo 7, tiene el honor de presentar su contribución nacional al informe que el Secretario General presentará en el octogésimo período de sesiones de la Asamblea para que los Estados Miembros sigan discutiendo la cuestión.

Introducción

La Federación de Rusia concede gran importancia a las cuestiones relativas a la aplicación de la inteligencia artificial (IA) en el ámbito militar y desea que se sigan examinando a fondo en foros internacionales especializados.

Consideramos que el Grupo de Expertos Gubernamentales sobre los Sistemas de Armas Autónomos Letales, creado por las Altas Partes Contratantes en la Convención sobre Ciertas Armas Convencionales, es el foro idóneo. Precisamente este Grupo está llamado a mantener un equilibrio razonable entre las preocupaciones humanitarias y los legítimos intereses de defensa de los Estados en relación con dichas armas y a tomar decisiones sobre la base del consenso. Su examen de las

aplicaciones militares de la inteligencia artificial es de amplio alcance, no se circunscribe a los sistemas de armas autónomos letales y aborda aspectos importantes (jurídicos, técnicos y militares) del uso de la tecnología con fines militares.

Observamos que este tema se está debatiendo en el marco de los sistemas vigentes de control de armas, desarme y no proliferación. El presente documento se centra en el análisis de los riesgos y oportunidades que la inteligencia artificial plantea a los Estados partes respecto del cumplimiento de las obligaciones que les incumben en virtud de los instrumentos jurídicos internacionales pertinentes.

Celebramos la disposición de los Estados Miembros a entablar en la Comisión de Desarme, en el marco del examen de las tecnologías emergentes en el contexto de la seguridad internacional, un debate sobre las aplicaciones militares de la inteligencia artificial. El objetivo de este intercambio de pareceres es acordar recomendaciones sobre aspectos de la IA “militar” que no se abordan en otros foros.

Durante los trabajos de los foros internacionales mencionados deberá prestarse especial atención a la elaboración de una terminología común, la aplicación del derecho internacional vigente, el control humano, la rendición de cuentas y los riesgos y oportunidades que plantea la tecnología.

Definición

En el derecho internacional vigente no hay una definición consensuada de los sistemas de armas y equipos militares basados en la IA, lo que hace difícil abordar la cuestión. Establecer un planteamiento común de trabajo sobre estas herramientas y, en general, la terminología asociada a la aplicación militar de esta tecnología permitirá tener una idea más clara del tema y de los enfoques con los que abordarlo.

Ese planteamiento de trabajo debería:

- a) Comprender una descripción de los tipos de sistemas de armas y equipos militares basados en la IA y las principales características específicas de su utilización;
- b) En vez de limitarse a la comprensión actual de estas herramientas, dar cuenta de cómo podrían evolucionar en el futuro;
- c) Ser universalmente aceptado por los distintos expertos (científicos, ingenieros, técnicos, militares, juristas y especialistas en ética);
- d) Estar concebido de manera que no limite el progreso tecnológico ni socave la investigación en robótica e IA con fines pacíficos;
- e) No definir los sistemas de armas y equipos militares basados en la IA únicamente describiendo sus funciones.

Hay que evitar clasificar estas herramientas como “malas” o “buenas”; es decir, que no deben clasificarse en función de las preferencias políticas de un grupo concreto de Estados.

Los sistemas militares que en la actualidad gozan de un alto grado de automatización no han de encasillarse en una categoría “especial” que requiera restricciones y prohibiciones urgentes. Gracias, precisamente, a ese grado de automatización, los sistemas en cuestión pueden funcionar de manera eficaz en situaciones de combate dinámicas y en diversos entornos y operar de manera suficientemente concreta y precisa para garantizar que cumplen los principios y normas del derecho internacional, incluido el derecho internacional humanitario.

Los sistemas de armas y equipos militares basados en la inteligencia artificial y el derecho internacional

En general, se acepta que el derecho internacional vigente, incluido el derecho internacional humanitario, se aplica plenamente a los sistemas de armas que utilizan la IA.

La Federación de Rusia considera que actualmente no existen motivos de peso para imponer nuevas restricciones o prohibiciones a los sistemas de armas que utilizan la IA, ni para actualizar o adaptar el derecho internacional, incluido el derecho internacional humanitario, para abordar este tipo de herramientas. Los debates dirigidos a acordar algún tipo de “normas de conducta” o reglas y principios de uso “responsable” de los sistemas de armas y equipos militares basados en la IA también son prematuros. El concepto de uso “responsable” de la IA que promueven los países occidentales se basa en criterios sumamente polémicos y ajenos al derecho internacional (incluido el derecho internacional humanitario), plantea numerosas interrogantes y no cuenta con el apoyo consensuado de la comunidad internacional.

Los principios de humanidad, los dictados de la conciencia pública y el componente de derechos humanos no pueden utilizarse como condición absoluta y única suficiente para imponer restricciones y prohibiciones a determinados tipos de armas y equipos militares. Las preocupaciones que planteen los sistemas de armas y equipos militares basados en la IA han de abordarse aplicando de buena fe de las normas jurídicas internacionales vigentes.

La estricta observancia de las normas y los principios del derecho internacional, incluido el derecho internacional humanitario, en situaciones de conflicto armado sigue siendo una de las prioridades de la Federación de Rusia. Las Fuerzas Armadas de la Federación de Rusia se atienen estrictamente a las normas del derecho internacional humanitario consagradas en los instrumentos jurídicos federales y departamentales. Las cuestiones relativas al cumplimiento del derecho internacional humanitario, también en lo tocante al uso de nuevos tipos de armas, se tratan en los reglamentos y en los programas de capacitación del personal militar de todas las categorías. En 2022 se aprobó un documento conceptual de las Fuerzas Armadas de la Federación de Rusia sobre el desarrollo y uso de sistemas de armas que utilizan IA.

La legislación rusa se atiene plenamente a las directrices sobre sistemas de armas asistidos con IA que los Estados partes en la Convención sobre Ciertas Armas Convencionales aprobaron por consenso en 2019. Consideramos que seguir intercambiando información sobre prácticas concretas de aplicación nacional de esas directrices es una forma de fomentar la confianza y aumentar la transparencia.

Control de los sistemas de armas y equipos militares que utilizan inteligencia artificial

Consideramos que el que los seres humanos deban tener control sobre el funcionamiento de los sistemas de armas y equipos militares basados en la IA es una limitación importante. Por lo tanto, el ciclo de control de estas herramientas debe permitir que un operador humano o un sistema de control de nivel superior intervenga en su modo de funcionamiento y pueda incluso para desactivarlas parcial o totalmente.

Para la Federación de Rusia, la responsabilidad sobre las decisiones de utilizar la fuerza siempre será humana. El control se ejerce en función del conjunto de la información disponible en el momento de tomar decisiones, pero las formas y métodos concretos de control humano han de dejarse a la discreción de los Estados y el control directo no tiene por qué ser la única opción.

El control de estos sistemas y equipos puede ejercerse en las formas siguientes:

- a) Incrementando su fiabilidad y tolerancia a fallos;
- b) Limitando los tipos de objetivos;
- c) Limitando el tiempo de utilización, la cobertura geográfica y el ámbito de aplicación;
- d) Realizando intervenciones rápidas y desactivándolos;
- e) Poniendo a prueba los sistemas en entornos operacionales realistas;
- f) Confiando la utilización (el control) de estas herramientas a personas que dominan los procedimientos de uso;
- g) Supervisando la fabricación de componentes particulares y de los dispositivos en su conjunto;
- h) Supervisando el desmantelamiento y la eliminación de componentes particulares y de los dispositivos en su conjunto.

Consideramos que no procede introducir en el debate, a propuesta de determinados Estados, los conceptos de “control humano significativo”, “formas y grados de implicación humana”, “control y evaluación humanos apropiados al contexto” y “previsibilidad, fiabilidad, trazabilidad y explicabilidad”, que suelen carecer de base jurídica y solo sirven para politizar los debates.

Responsabilidad

La Federación de Rusia considera que los Estados y las personas (incluidos los desarrolladores y fabricantes) son responsables en todo momento, en virtud del derecho internacional, de sus decisiones de desarrollar y utilizar sistemas de armas y equipos militares basados en la IA. La responsabilidad del uso de estas herramientas recae en el funcionario que les asigna una tarea y da la orden de utilizarlos. Para utilizar sistemas de armas y equipos militares basados en la IA, dicho funcionario deberá poseer los conocimientos y aptitudes necesarios en relación con su funcionamiento y operación, y deberá ser responsable de tomar la decisión sobre la conveniencia de su uso y de planificar las formas y medios de su utilización.

Oportunidades y limitaciones de los sistemas de armas y equipos militares que utilizan la inteligencia artificial

Se sabe que los sistemas de armas y equipos militares que utilizan IA pueden realizar tareas con más eficacia y menos probabilidades de error que los operadores humanos. Este tipo de herramientas permiten en particular reducir sustancialmente las consecuencias negativas —según el derecho internacional, incluido el derecho internacional humanitario— asociadas a errores de los operadores, a su estado mental o físico o a sus creencias morales, religiosas o éticas. Su uso puede permitir ataques más precisos de instalaciones militares y contribuir a reducir el riesgo de impactos involuntarios en civiles y bienes de carácter civil.

Una evaluación de los riesgos potenciales relacionados con el uso de sistemas de armas y equipos militares basados en la IA y las medidas para mitigarlos deberían formar parte del proceso de diseño, desarrollo, prueba y despliegue de nuevas tecnologías en cualquier tipo de sistema militar.

Los riesgos asociados a estas herramientas podrían minimizarse de las formas siguientes:

- a) Garantizando una gestión eficaz del ciclo de vida;

- b) Realizando pruebas exhaustivas en todas las fases del ciclo de vida, incluso en entornos casi reales;
- c) Incrementando su fiabilidad y tolerancia a fallos;
- d) Fijando criterios de preparación;
- e) Garantizando la máxima protección contra accesos no autorizados;
- f) Impartiendo formación a los operadores;
- g) Primando el uso de tecnologías de IA para reunir y procesar información con la que facilitar la toma de decisiones en el ámbito militar;
- h) Facilitando la supervisión continua por parte del operador de las operaciones de dichos sistemas y permitiendo la finalización de emergencia de una misión de combate si el operador así lo ordena;
- i) Evitando que estas herramientas caigan en manos de agentes no estatales, que podrían utilizarlos con fines ilegales.

Estas medidas pueden adoptarse en todas las etapas del ciclo de vida (desarrollo, producción, utilización, eliminación) de las armas y los equipos militares y especiales.

Próximos pasos

A nuestro juicio, convendría que los Estados siguieran examinando las cuestiones relacionadas con la aplicación militar de la IA en el seno del Grupo de Expertos Gubernamentales sobre los Sistemas de Armas Autónomos Letales, que es el foro internacional más indicado para hacer tal examen en el contexto de los sistemas vigentes de control de armamentos, desarme y no proliferación, así como en la Comisión de Desarme. Asimismo, las deliberaciones que tengan lugar en un foro no deben duplicar las que se lleven a cabo en otros.

Nos oponemos a la fragmentación de iniciativas en este ámbito. Parece contraproducente trasladar la cuestión del uso de la IA con fines militares a cualquier otra plataforma internacional, establecer nuevos foros que la examinen o sometan a debate en un marco restringido en el que no participen la inmensa mayoría de los Estados Miembros de las Naciones Unidas (incluidos los principales desarrolladores de sistemas de armas basados en la IA, como la Federación de Rusia).

En concreto, no es constructivo abordar estos temas en cumbres no inclusivas “sobre el uso responsable de la IA con fines militares” organizadas por un grupo de Estados occidentales, como tampoco lo son las cumbres sobre la IA en general. Esos actos y sus documentos finales no tienen en cuenta las opiniones de todos los interesados y no pueden servir de base para un trabajo ulterior que refleje un entendimiento común del tema. Se trata de actividades divisivas que no conducen a aunar esfuerzos en este ámbito.

Los intentos de “consolidar” enfoques unilaterales sobre estas cuestiones en foros alternativos, como las “cumbres” antes mencionadas, al margen de los foros multilaterales especializados, tendrán consecuencias sumamente negativas, ya que podrían socavar gravemente la labor constructiva e inclusiva que está realizándose al respecto y fragmentar las iniciativas encaminadas a establecer premisas y recomendaciones comunes en este ámbito.

Consideramos que, en los foros internacionales mencionados, los debates deberían centrarse sobre todo en acordar términos y planteamientos concretos comunes respecto de la aplicación del derecho internacional vigente, incluido el derecho internacional humanitario, a los sistemas de armas y equipos militares que

utilizan la IA, en garantizar que esas herramientas se mantengan bajo control humano y en abordar los riesgos y oportunidades que conlleva esta tecnología.

La Federación de Rusia solicita al Secretario General que, en su informe sustantivo presentado en cumplimiento de lo dispuesto en el párrafo 7 de la resolución [79/239](#) de la Asamblea General, tenga en cuenta las propuestas arriba planteadas y que incluya el presente documento en el anexo de ese informe.

Finlandia

[Original: inglés]
[11 de abril de 2025]

Finlandia se complace en presentar sus opiniones sobre la resolución [79/239](#) de la Asamblea General, relativa a la inteligencia artificial (IA) en el ámbito militar y a sus implicaciones para la seguridad internacional, que se aprobó el 24 de diciembre de 2024 y en la que la Asamblea solicita al Secretario General que recabe las opiniones de los Estados Miembros sobre “las oportunidades y los retos que plantea para la paz y la seguridad internacionales la aplicación de la inteligencia artificial en el ámbito militar, con atención específica a los ámbitos distintos de los sistemas de armas autónomos letales”.

La aprobación de principios o normativas internacionales sobre la aplicación de la inteligencia artificial en el ámbito militar es fundamental para garantizar el cumplimiento del derecho internacional, aumentar la seguridad y reducir los riesgos potenciales de conflictos. Al mismo tiempo, es necesario permitir el desarrollo de capacidades nacionales de defensa que sí se ajusten al derecho internacional. Finlandia se ha comprometido a desarrollar, desplegar y utilizar capacidades de IA en el ámbito militar de forma responsable, de conformidad con el derecho internacional, en particular el derecho internacional humanitario, y de manera que no socave la paz, la seguridad y la estabilidad internacionales, al tiempo que prosigue sus esfuerzos de investigación, desarrollo, experimentación e innovación en el ámbito de la tecnología de la IA.

Resulta cada vez más importante determinar las implicaciones de las tecnologías disruptivas en la política exterior, la seguridad y la defensa y crear medios para abordarlas. Finlandia participa activamente en los debates mundiales sobre la regulación de la tecnología, en los que defiende los derechos fundamentales y humanos y aborda los riesgos relacionados con el desarrollo y la aplicación de la IA y las políticas pertinentes.

Además de determinar los riesgos que plantean las tecnologías disruptivas, también es importante reconocer las oportunidades que ofrecen para la seguridad, el desarrollo de las capacidades de defensa, el crecimiento económico, la productividad, el desarrollo sostenible, la competencia tecnológica y las inversiones sectoriales.

Oportunidades

Las tecnologías disruptivas ofrecen importantes oportunidades para hacer avanzar diversos sectores, impulsar la transición limpia, fomentar el crecimiento económico sostenible y mejorar la eficiencia y la productividad. También tienen el potencial de mejorar la seguridad, la educación, el bienestar y la salud en todo el mundo.

La IA y otras tecnologías emergentes ofrecen oportunidades para avanzar en las capacidades de defensa y contribuyen de manera determinante a configurar los campos de batalla y los medios y métodos de guerra del futuro. Los avances

tecnológicos permiten reunir información y procesar datos con mayor eficacia, mejorar la conciencia situacional, tomar decisiones más rápido e intervenir de manera más precisa y amplia. En las formas contemporáneas de hacer la guerra tienen cada vez más importancia los sistemas teledirigidos y los sistemas autónomos no tripulados, que cambiarán el futuro de la guerra, las operaciones y el campo de batalla. Anticiparse a los avances tecnológicos, integrar las tecnologías emergentes en los sistemas de defensa y aprovechar novedades inesperadas será cada vez más importante, a medida que el ritmo del desarrollo tecnológico se acelere en el futuro. La ventaja tecnológica también puede compensar la inferioridad numérica.

Desafíos

Al mismo tiempo, es importante forjar un amplio consenso sobre lo que se entiende por amenazas a la seguridad, posibles usos indebidos, cuestiones de derechos humanos e interdependencias en el desarrollo de tecnologías disruptivas, como la IA. A medida que se desarrolle, estas tecnologías plantearán nuevos retos, sobre todo para los sectores de la defensa y la seguridad. El desarrollo de la IA hace que los ciberataques, las actividades de influencia informativa y, uno de sus instrumentos, la desinformación, estén mejor orientados y sean más eficaces. Además, la IA ya se está utilizando para influir en procesos electorales. En estas circunstancias, también debe prestarse mayor atención a la seguridad de la información confidencial.

El derecho internacional vigente, en particular la Carta de las Naciones Unidas, el derecho internacional de los derechos humanos y el derecho internacional humanitario, se aplica plenamente al ciberespacio. El respeto y la adhesión al marco de comportamiento responsable de los Estados en el ciberespacio siguen siendo esenciales para mantener la paz, la seguridad y la estabilidad internacionales. El desarrollo tecnológico plantea nuevas interrogantes sobre, por ejemplo, el ciberentorno, el uso de la IA, las nuevas tecnologías armamentistas y la explotación de materias primas críticas. Las actividades de influencia híbrida pueden entrañar prácticas que entorpezcan la rendición de cuentas en virtud del derecho internacional. Finlandia aboga por tener muy en cuenta los derechos fundamentales y humanos y los riesgos conexos a la hora de desarrollar y aplicar la IA y de establecer normas para su regulación. Es importante fijar principios, criterios y normas, políticas y marcos nacionales a fin de garantizar una aplicación responsable de la IA en el ámbito militar, de conformidad con el derecho internacional.

El desarrollo tecnológico ha aportado a actores hostiles nuevas posibilidades de realizar actividades híbridas de influencia que por su intensidad no llegan a la categoría de conflicto abierto. Las ciberoperaciones hostiles se han convertido en una constante de la política del poder y están entre las herramientas que utilizan los agentes estatales para ejercer influencia. También se realizan operaciones cibernéticas, híbridas y de información en condiciones normales, lo que puede difuminar los límites entre la guerra y la paz. A pesar de que la tecnología tiene un papel cada vez mayor en la guerra, las capacidades bélicas convencionales siguen siendo importantes, sobre todo en conflictos amplios y duraderos.

Muchos países están afrontando intensas actividades de influencia informativa que también utilizan la IA. El uso perjudicial de la información ha llegado a ser un componente habitual de las campañas de influencia de amplio espectro y la competencia en el terreno de la información ha aumentado.

La evolución de las infraestructuras y la tecnología, así como el número cada vez mayor de usuarios, incrementan las posibilidades de llevar a cabo acciones hostiles en el ciberespacio. Muchos países afrontan constantemente actividades de captación de inteligencia en redes de información, ciberspying y ciberataques de actores hostiles, que también tratan de influir materialmente en infraestructuras

críticas. Están aumentando las actividades hostiles orquestadas, no ya por actores estatales, sino por actores no estatales que persiguen fines políticos o siguen instrucciones de Estados.

Francia

[Original: francés]
[11 de abril de 2025]

I. La inteligencia artificial en el ámbito militar y sus implicaciones para la paz y la seguridad internacionales

Oportunidades que pueden aprovecharse

Facilitar la planificación y la toma de decisiones. Las fuerzas armadas francesas trabajan en la explotación de sus bases de datos sobre sucesos relacionados con municiones y explosivos para elaborar herramientas con las que predecir posibles amenazas en zonas predeterminadas.

Apoyar a las personas. El sistema de inteligencia artificial (IA) para la capacitación de tripulantes de vuelo (“IA FPN”) se está utilizando para mejorar la formación de pilotos franceses mediante el análisis de datos de vuelos o simulaciones. La IA también puede ayudar a personas expuestas a grandes cantidades de datos, como hace el sistema “oreille d’or”, que procesa en masa datos acústicos para que la atención de los operadores franceses se centre exclusivamente en las señales con valor agregado.

Contrarrestar nuestras vulnerabilidades en el ámbito de las tecnologías de la información y las comunicaciones. Las tecnologías de IA pueden aplicarse a la ciberseguridad y para contrarrestar la proliferación de información falsa. Las fuerzas armadas francesas disponen de sistemas de detección de ultrafalsificaciones (*deepfakes*).

Promover la aplicación del derecho internacional humanitario y la protección de las personas y los bienes protegidos. La IA puede facilitar la aplicación de los principios de distinción, proporcionalidad y precaución, entre otros principios cardinales del derecho internacional humanitario. También puede ser útil para proteger a las personas al facilitar la remoción de minas terrestres mediante drones dotados de sensores conectados a dispositivos de IA.

Reforzar el control de armamentos. La IA puede servir para controlar y detectar mejor los lanzamientos clandestinos, la reubicación de instalaciones de fabricación de armas o los ensayos de armas químicas y biológicas. También puede utilizarse para controlar las exportaciones de armas, ya que sirve para mejorar la trazabilidad.

Reforzar la prevención, el mantenimiento y la consolidación de la paz. La IA permite realizar operaciones de mantenimiento de la paz mejor adaptadas y, por tanto, más eficaces. El sistema de traducción instantánea “Resistance”, propuesto por las fuerzas armadas francesas, se concibió para comunicarse con la población local cuando no se dispone de conexión a Internet, lo que permite luchar contra la desinformación.

Riesgos que deben mitigarse

Riesgos inherentes a la tecnología. Las técnicas de aprendizaje entrañan diferentes riesgos de sesgo: sesgos involuntarios; sesgos voluntarios; sesgos de reconstitución de datos especialmente sensibles; resultados confusos o poco

inteligibles. También se plantea la cuestión del consumo exponencial de recursos energéticos.

Agravamiento de los riesgos para la seguridad y la estabilidad internacional. En manos equivocadas, la IA puede agravar ciertos riesgos para la seguridad y la estabilidad internacionales (hipótesis de escalada, carrera armamentista, proliferación hacia actores no estatales, expansión de las operaciones de influencia y las acciones hostiles en el ciberspace), para los que habrá que adaptar las medidas de mitigación. Los riesgos de desresponsabilización inducidos por la dependencia de las tecnologías exigen garantizar la responsabilidad humana.

II. Principios y medidas clave para una inteligencia artificial responsable en todo el ciclo de vida

Desarrollar una inteligencia artificial que se atenga al derecho internacional humanitario

Adaptar el examen de legalidad. Aunque este examen es plenamente aplicable a la IA militar, las modalidades precisas deberán adaptarse a las especificidades de esta tecnología.

Hacer los controles necesarios Estos controles deben realizarse en las distintas fases del ciclo de vida de un sistema de armas según se necesite y cuando en un dispositivo se introduzcan innovaciones o nuevos componentes que puedan alterar sustancialmente los efectos previstos.

Desarrollar un sistema de inteligencia artificial fiable y seguro

Evaluuar, calificar y certificar los sistemas. Estos sistemas deben evaluarse y calificarse debidamente (según el grado de criticalidad de las funciones) mediante un análisis de riesgos durante la etapa de diseño. Deben asociarse a usos precisos. Ha de preverse la renovación periódica de estos controles en función de la importancia de los aspectos controlados.

Basarse en datos controlados soberanos. Conviene establecer medidas para afrontar riesgos de protección de los datos y salvaguardas adecuadas.

Corregir y reentrenar los sistemas. Es importante registrar y caracterizar los errores detectados (al probarlos o trabajar con ellos), transmitir a los operadores las lecciones aprendidas y comprobar continuamente que el sistema se ajusta a nuestras obligaciones internacionales.

Someter la inteligencia artificial al control humano adecuado y a una cadena de mando responsable

Asegurarse de que las decisiones y actuaciones son conformes a derecho. El operador o mando militar ha de valerse de su propio discernimiento para comprobar que los resultados propuestos se ajustan a las órdenes recibidas y a las obligaciones legales.

Garantizar la responsabilidad humana. La responsabilidad humana en la concepción, el despliegue y el uso de tecnologías de IA es un principio irrevocable, que exige formalizar las cadenas de responsabilidad de quienes ejercen funciones de mando, control y ejecución.

Adaptar el control humano. Analizar y caracterizar el control humano adecuado sin que ello entrañe restringir las capacidades de los sistemas que funcionan con tecnologías de IA es una labor compleja en la que han de tenerse en cuenta diversos factores humanos, técnicos y contextuales.

Capacitar a los mandos y el personal militares para que dominen estos sistemas. Debe establecerse una etapa de capacitación y práctica sobre los sistemas antes de ponerlos en uso para que el personal conozca sus ventajas y riesgos.

Desarrollar una inteligencia artificial sostenible

Proteger la investigación. Los programas de investigación deben tener propósitos y metas abiertos y no ha de imponérseles *a priori* restricciones excesivamente amplias.

Someter la investigación a una reflexión ética. Francia ha creado un órgano permanente de reflexión sobre las cuestiones éticas que plantean las nuevas tecnologías en el ámbito de la defensa: el Comité de ética de la defensa.

Desarrollar una inteligencia artificial sobria. Fomentar un funcionamiento sobrio entraña reflexionar sobre el uso de la IA y mejorar la resiliencia y la sostenibilidad de los sistemas manteniendo los costos bajo control.

III. Un proceso específico para establecer un sistema de gobernanza mundial que permita poner en práctica los principios de la inteligencia artificial responsable

Un proceso universal inclusivo. Los debates deben reunir a todas las partes interesadas: los Estados (en concreto, la participación activa de los Estados que desarrollan y utilizan estos sistemas es absolutamente indispensable; con este imperativo, el proceso de toma de decisiones deberá tener en cuenta las distintas posiciones y adoptar normas que vayan en ese sentido en aras del consenso), pero también las empresas del sector, el ámbito científico y académico y la sociedad civil, para evitar debates desconectados de la realidad y preservar la innovación. La Primera Comisión de la Asamblea General puede ser un foro adecuado.

Una arquitectura de gobernanza racionalizada y coherente. Un marco único debería permitir racionalizar los esfuerzos para ganar en eficacia y reforzar la incidencia de los resultados. Será indispensable velar por la complementariedad con los debates del Grupo de Expertos Gubernamentales sobre las Tecnologías Emergentes en el Ámbito de los Sistemas de Armas Autónomos Letales, que debería poder proseguir sus trabajos después de 2026, bajo un nuevo mandato.

Un proceso de orientación operacional centrado en las cuestiones de interés en el ámbito militar. La gobernanza debe atenerse al corpus jurídico aplicable a los conflictos armados y, por consiguiente, en el derecho internacional humanitario en primer lugar. La prioridad de cualquier proceso internacional debe ser garantizar el respeto de las normas jurídicas vigentes, para lo que deberá tratarse del establecimiento de principios rectores y de los recursos que permitan a los Estados aplicarlos (facilitando el intercambio de mejores prácticas y fomentando la cooperación y la asistencia internacionales por medios que se adapten a los asuntos militares), al tiempo que se promueven medidas adecuadas de fomento de la confianza y de reducción del riesgo.

Grecia

[Original: inglés]
[10 de abril de 2025]

La integración de la inteligencia artificial (IA) en el sector de la defensa ha tenido una influencia determinante en los medios y los métodos de realización de operaciones militares. Las aplicaciones militares de la IA han aportado grandes beneficios operacionales, como una toma de decisiones más rápida, una mejor

detección y predicción de amenazas, el conocimiento y la evaluación de la situación en tiempo real, la optimización de la asignación y la planificación de recursos, el refuerzo de la logística, la ampliación de las capacidades humanas en tareas complejas y el procesamiento eficaz de grandes volúmenes de datos de inteligencia.

Pero, a pesar de estos avances, es esencial reconocer que el progreso tecnológico también acarrea retos complejos y multidimensionales que deben examinarse cuidadosamente para asegurarse de que no socavan la paz, la seguridad y la estabilidad regionales y mundiales.

A este respecto, una de las principales preocupaciones de Grecia es el uso de sistemas militares con capacidades de aprendizaje automático, que plantea diversos retos (como la transparencia y la explicabilidad), ya que los modelos complejos pueden funcionar como “cajas negras” con procesos de toma de decisiones imprecisos, sobre todo habida cuenta de que los entornos de combate cambian constantemente.

Además, el posible uso de la IA generativa en equipos militares introduce un factor importante de complejidad e incertidumbre, ya que esos sistemas podrían generar de forma autónoma soluciones novedosas y adaptarse a las condiciones cambiantes del campo de batalla gracias a un proceso constante de análisis y aprendizaje a partir de datos nuevos, una capacidad que Grecia considera sumamente preocupante. Para hacer frente a esos retos, es crucial imponer unos límites y restricciones operacionales claros al uso de esos sistemas a fin de evitar comportamientos no deseados.

Dadas las circunstancias, uno de los problemas más alarmantes que plantea el uso de la IA en contextos militares es su integración en sistemas de mando y control y de apoyo para la toma de decisiones para el uso de armas nucleares. La posibilidad de delegar en sistemas que funcionan con IA decisiones relativas a la disuisión nuclear o incluso de activar los protocolos de uso pertinentes ha de examinarse con sumo cuidado para garantizar que tales decisiones están sujetas a supervisión e implicación humanas y que se establecen las salvaguardas de ciberseguridad adecuadas para evitar escaladas indeseadas.

En el complejo entorno geopolítico actual, resulta igualmente preocupante que los Estados se esfuerzen por mantener la superioridad militar, lo que podría conducir a una carrera armamentista marcada por la falta de transparencia y la sospecha mutua. Esa competición puede agravar la inestabilidad geopolítica y plantear importantes amenazas para la seguridad mundial si se altera el equilibrio de poderes y la brecha tecnológica entre Estados avanzados y en desarrollo se vuelve aún mayor.

Además, el creciente desarrollo y despliegue de capacidades de IA por parte de fuerzas armadas podría rebajar el umbral de los conflictos armados. La aceleración de la toma de decisiones y la creciente dependencia de sistemas no tripulados en los teatros de operaciones aumentan el riesgo de escalada involuntaria, ya que el elemento humano en el campo de batalla se sustituye cada vez más por sistemas no tripulados.

En este contexto, otro parámetro que requiere la consideración adecuada es la proliferación y desvío de capacidades de IA hacia Estados que hacen caso omiso del orden internacional sujeto a normas y hacia actores no estatales, como las organizaciones terroristas. A medida que aumenta el acceso a las tecnologías de IA, crece sustancialmente el riesgo de que actores de ese tipo puedan adquirirlas y desplegarlas para fines de desestabilización, lo que supondría un reto aún mayor para la seguridad internacional.

Las aplicaciones militares de IA también generan riesgos y problemas en relación con operaciones psicológicas y de información errónea, ya que permiten generar grandes volúmenes de información falsa, ultrafalsificaciones y datos tergiversados a fin de engañar a los ciudadanos y desestabilizar las instituciones. Las cuentas automatizadas (bots) y los algoritmos de propaganda selectiva refuerzan las operaciones psicológicas al influir en la opinión pública y los procesos electorales y generar tensiones sociales, llegando incluso a minar la confianza de la población en operaciones de mantenimiento de la paz con campañas de desinformación. Sesgos sociales, como los relativos al género, la edad, la raza y la discapacidad, también generan preocupación, y es esencial hacer evaluaciones de riesgos y aplicar medidas de mitigación que eviten sesgos y discriminaciones involuntarios en los algoritmos.

Además, las aplicaciones de la IA en el ámbito de la ciberseguridad pueden utilizarse tanto para proteger infraestructuras críticas como con fines maliciosos, como los ciberataques y la interceptación de datos. Las amenazas híbridas que combinan operaciones militares tradicionales con tácticas ofensivas de inteligencia requieren una mayor vigilancia y coordinación entre actores estatales e internacionales para evitar escaladas y preservar la paz y la seguridad regionales e internacionales.

Habida cuenta de lo expuesto, Grecia apoya firmemente las iniciativas internacionales encaminadas a garantizar el uso responsable de la IA en el ámbito militar, ya que, a pesar de los retos que acaban de describirse, puede reforzar la aplicación del derecho internacional humanitario y contribuir a la protección de los civiles al mejorar la precisión de los objetivos, fortalecer la vigilancia y optimizar la asistencia humanitaria.

Con ese espíritu, en colaboración con Francia y la República de Corea y con el valioso apoyo de Armenia, Italia y el Reino de los Países Bajos, Grecia celebró el 4 de abril de 2025, con arreglo a la fórmula Arria, una reunión del Consejo de Seguridad sobre el uso seguro, inclusivo y fiable de la inteligencia artificial para el mantenimiento de la paz y la seguridad internacionales. En ella se aportaron valiosas perspectivas sobre las formas en que las Naciones Unidas pueden contribuir al mantenimiento de la paz y la seguridad internacionales, en particular la regulación, la no proliferación y la prevención del desvío de capacidades militares de IA, el fortalecimiento del estado de derecho, los valores democráticos, la cohesión social y el desarrollo económico.

Además, en el marco de su actividad internacional, Grecia ha apoyado las declaraciones conjuntas dimanadas de las dos ediciones de la Cumbre sobre la Inteligencia Artificial Responsable en el Ámbito Militar, celebradas en La Haya (15 y 16 de febrero de 2023) y Seúl (9 y 10 de septiembre de 2024), que tratan de las medidas de desarrollo y uso responsables de la IA en el ámbito militar. También ha respaldado la Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar, promovida por los Estados Unidos de América, y la Declaración de París sobre el Mantenimiento del Control Humano en los Sistemas de Armas que Funcionan con IA.

Además, Grecia ha creado una comisión consultiva de alto nivel sobre IA¹ para elaborar una estrategia nacional global en este ámbito, junto con las estructuras necesarias dentro del Ministerio de Defensa Nacional, para abordar los retos

¹ En el estudio de referencia de esta Comisión, titulado “Un plan para transformar Grecia con IA”, se plantean principios rectores y proyectos emblemáticos para impulsar el desarrollo de la inteligencia artificial en Grecia y se establecen prioridades, como preservar y mejorar la democracia, adaptarse al cambio climático y mitigar sus efectos y reforzar la seguridad.

tecnológicos, jurídicos, éticos y políticos que plantean las aplicaciones de la IA y la autonomía en el ámbito militar.

Por último, para contribuir de forma constructiva al diálogo internacional sobre el uso responsable de la IA en el ámbito militar, Grecia está organizando una conferencia internacional sobre los conflictos armados y la gestión de crisis en la era de la IA, que se celebrará en Atenas los días 22 y 23 de mayo de 2025.

India

[Original: inglés]
[1 de abril de 2025]

La inteligencia artificial (IA) es una tecnología transformadora que está teniendo importantes repercusiones en todos los aspectos de la vida humana. Se está desarrollando a una escala y velocidad sin precedentes y se está adoptando y desplegando rápidamente para una serie de aplicaciones. La IA puede tener efectos transformadores en la reducción de la pobreza y la mejora de la vida de las personas, algo especialmente importante para países en desarrollo como la India.

Hay que hacer un trabajo colectivo mundial para establecer una gobernanza y unas normas para la IA que permitan preservar nuestros valores compartidos, tratar los riesgos y fomentar la confianza. Esta gobernanza y esas normas deben tener en cuenta la profunda interdependencia allende las fronteras; promover la innovación; desplegarse para el bien general; y promover el acceso y la equidad para que los beneficios de la IA estén al alcance de todos, especialmente de los países del Sur Global. La India aboga por unos debates abiertos sobre innovación y gobernanza.

Los debates sobre la IA militar deben anclarse en la realidad militar; y la IA está integrándose con rapidez en las doctrinas y operaciones militares. Diversos conflictos en todo el mundo han puesto de manifiesto los riesgos y las oportunidades que se derivan del uso cada vez mayor de estas tecnologías.

El desarrollo, despliegue y uso de la IA en el ámbito militar plantea retos éticos, legales y de seguridad. Sin restarles importancia, la India comparte la opinión de que la IA puede mejorar el cumplimiento del derecho internacional humanitario.

La India respalda los esfuerzos colectivos mundiales por regular adecuadamente el desarrollo, el despliegue y el uso de la IA en el ámbito militar, que deben tener en cuenta las consideraciones jurídicas y éticas y detectar y mitigar los riesgos que esta tecnología plantea en el ámbito militar.

Todo esfuerzo colectivo para regular adecuadamente la IA ha de centrarse en sus aplicaciones y usos, no en la propia tecnología y sus componentes. Es necesario evitar la estigmatización de estas tecnologías y no restringir su uso para fines de desarrollo.

La IA debe utilizarse legalmente en el ámbito militar, conforme al derecho inmanente de legítima defensa individual o colectiva que reconoce el derecho internacional. El derecho internacional humanitario sigue siendo plenamente aplicable a la IA en el ámbito militar. Los principios cardinales del derecho internacional humanitario, a saber, la distinción, la proporcionalidad y la precaución, se aplican a todos los medios y métodos de guerra pasados, presentes y futuros.

Es indispensable que el uso de la IA en el ámbito militar esté sujeto a discernimiento y supervisión humanos para mitigar los riesgos y garantizar el cumplimiento del derecho internacional humanitario.

Toda labor colectiva o regulación adecuada en relación con la IA en el ámbito militar debe tener en cuenta las obligaciones jurídicas vigentes y respetar la jurisdicción y competencia nacionales, así como las capacidades nacionales pertinentes.

La India está comprometida con el uso responsable de la IA en el ámbito militar.

La India está elaborando un marco para evaluar la fiabilidad de la IA en el sector de la defensa a fin de abordar los complejos retos que plantean esta tecnología moderna. Se centra en cinco principios fundamentales: a) fiabilidad y solidez; b) seguridad; c) transparencia; d) equidad; y e) privacidad. Estos principios ofrecen una base para examinar cómo regular adecuadamente el desarrollo, el despliegue y el uso de la IA en el ámbito militar.

Indonesia

[Original: inglés]
[11 de abril de 2025]

Indonesia acoge con satisfacción el debate sobre las oportunidades y los retos para la paz y la seguridad internacionales que plantea la aplicación de la inteligencia artificial (IA) en el ámbito militar, con especial atención a los ámbitos distintos de las armas autónomas letales, según lo dispuesto en los párrafos 7 y 8 de la resolución [79/239](#) de la Asamblea General.

Como la IA en el ámbito militar comprende una amplia gama de sistemas y aplicaciones, una deliberación multilateral inclusiva sobre el tema en las Naciones Unidas debería abarcar, además de las capacidades cinéticas (como los sistemas de armas autónomos letales), las no cinéticas, que en las funciones militares pueden ser de agresión (por ejemplo, los sistemas autónomos de ciberguerra, las capacidades adaptativas de interferencia de radares o guerra electrónica) o de apoyo (por ejemplo, las de logística, evacuación médica o vigilancia táctica), así como otras capacidades que pueden afectar directamente al equilibrio estratégico, como las de detección mejorada (por ejemplo, satelital o antisubmarina), inteligencia o planificación de la guerra.

Indonesia sigue firmemente comprometida con el mantenimiento de la paz y la seguridad internacionales, como se indica en el preámbulo de su Constitución. Guiada por este compromiso, Indonesia cree que el uso de la IA en el ámbito militar debe servir para promover la paz, la seguridad y los Objetivos de Desarrollo Sostenible. La IA debe ser una fuerza de paz y seguridad, no un factor de inseguridad, conflictos o rivalidad estratégica.

Aunque la IA en sí misma no es un arma, Indonesia es consciente de que sirve tanto para multiplicar fuerzas como para amplificar amenazas y puede generar importantes beneficios y graves riesgos para la paz y la seguridad internacionales. El uso de la IA en el ámbito militar plantea diversas cuestiones éticas, jurídicas, morales y técnicas, que deben ser cuidadosamente examinadas y consideradas en relación con el cumplimiento del derecho internacional, en particular el derecho internacional humanitario y de los derechos humanos.

Por un lado, se considera que la IA ofrece una amplia gama de posibilidades (mejorar el procesamiento de datos; incrementar la eficacia, la precisión y la exactitud operacionales; y mejorar potencialmente el cumplimiento del derecho internacional humanitario, por ejemplo al reforzar la evaluación de la proporcionalidad y de las precauciones tomadas para reducir el daño a civiles) y puede ampliar las capacidades

de inteligencia, vigilancia y reconocimiento, reforzar la logística y la planificación y mejorar la gestión del personal.

Por otro, entraña diversos riesgos y consecuencias (como desencadenar carreras armamentistas, proliferar entre actores no estatales, propiciar usos indebidos delictivos e irresponsables, agravar los desequilibrios de poder militar a través de la superioridad tecnológica e incrementar la inestabilidad, los errores de cálculo, las tensiones y las ambigüedades jurídicas), así como riesgos técnicos, como cibervulnerabilidades, fallos de funcionamiento de los sistemas, datos sesgados, identificaciones erróneas de objetivos y otras incertidumbres operacionales.

Indonesia está especialmente preocupada por los riesgos existenciales de la posible integración de la IA en los sistemas nucleares de mando, control y comunicación. Indonesia reafirma su posición de principio de que utilizar armas nucleares o amenazar con su uso viola el derecho internacional y de que deben aplicarse con urgencia y determinación medidas que preserven y refuerzen las normas contra las armas nucleares. La introducción de la IA en los sistemas de armas nucleares agrava los riesgos existenciales asociados al uso de armas nucleares, ya sea intencional, fortuito o accidental, y aumenta los peligros nucleares, lo que constituye una amenaza para la seguridad de todas las naciones. Indonesia insta a todos los Estados poseedores de armas nucleares a reevaluar su dependencia de ellas y a reafirmar nuestro compromiso con un mundo sin armas nucleares. Hasta que no se eliminen todas las armas nucleares, los Estados poseedores deben mantener tales armas y sus sistemas vectores bajo el control, la responsabilidad y la rendición de cuentas de personas en el contexto del desarrollo de la IA.

A la luz de estas consideraciones, Indonesia insta a adoptar un enfoque preventivo para hacer frente a los retos que plantea el uso de la IA en el ámbito militar y subraya que el desarrollo, la aplicación y el uso de la IA en el ámbito militar deben gestionarse de modo que se aprovechen sus beneficios y se mitiguen sus riesgos. Esta gestión debe estar al servicio de la paz, la seguridad y la prosperidad colectivas de todas las naciones. Sobre esa base, Indonesia presenta los siguientes puntos clave:

En primer lugar, Indonesia afirma que el derecho internacional debe respetarse durante todo el ciclo de vida de las tecnologías de IA, lo que incluye la Carta de las Naciones Unidas, el derecho internacional humanitario, el derecho internacional de los derechos humanos y los tratados sobre desarme y no proliferación. Los Estados deben hacer exámenes legales en todas las fases, desde la adquisición hasta la evaluación. Los Estados deben garantizar la rendición de cuentas por el desarrollo y la aplicación de capacidades de IA en el ámbito militar y asegurarse de que es legal utilizar aplicaciones de la IA para la conducción de guerras u hostilidades. Si no hay leyes que regulen el uso de la IA en el ámbito militar, este habrá de regirse por las leyes de la humanidad y los dictados de la conciencia pública.

Más allá del derecho internacional, las consideraciones éticas deben complementar los marcos jurídicos para orientar la gobernanza del uso de la IA en el ámbito militar. Al desarrollar y aplicar capacidades de IA, han de promoverse principios como la trazabilidad, la rendición de cuentas, la responsabilidad, la explicabilidad, la humanidad, la transparencia, la equidad y la justicia.

En segundo lugar, Indonesia subraya función esencial del elemento humano en lo que respecta a la rendición de cuentas y la responsabilidad a todos los niveles, ya sea estatal, empresarial o individual, en la concepción, el desarrollo, el despliegue y el uso de la IA en el ámbito militar.

El desarrollo, la aplicación y el uso de la IA en el ámbito militar deben seguir centrándose en el ser humano y utilizándose en interés de la humanidad. El control humano efectivo y determinante, en particular de las decisiones que entrañen el uso

de la fuerza, debe preservarse y reforzarse con medidas de capacitación. Las decisiones críticas deben estar sujetas al juicio, la intervención, la supervisión y el control humanos. Además, aunque el “control humano determinante” se ha ido aceptando cada vez más en la gestión del uso de la IA en el ámbito militar, Indonesia considera que este concepto aún tiene que satisfacer los criterios jurídicos, morales, técnicos y normativos asociadas a ese uso. Es necesario llegar a un acuerdo sobre lo que implica en la práctica un control humano “determinante”.

Aunque la gobernanza de la IA regulará principalmente la conducta de los Estados, también debe ocuparse de las partes civiles, en particular las empresas tecnológicas implicadas en el uso de la IA en el ámbito militar. Los Estados deben asegurarse de que el sector privado cumple con el derecho internacional y las normas éticas sin dejar de apoyar el crecimiento del ecosistema de innovación de la IA. Los investigadores y las empresas tienen la obligación de velar por que sus tecnologías de IA sean fiables, seguras, entrañen rendición de cuentas y estén bajo un control humano responsable. También están obligados a supervisar, dar a conocer y tratar los riesgos que entrañan sus productos.

En tercer lugar, Indonesia subraya la necesidad acuciante de marcos de gobernanza legales y reglamentarios multilaterales, inclusivos y exhaustivos, que deben reflejar los intereses de todos los Estados, independientemente de su nivel de desarrollo de capacidades de IA. Todos los Estados deben poder participar en pie de igualdad en la configuración de las reglas y normas que regulen el uso de la IA en el ámbito militar a fin de garantizar una representación justa y fomentar la confianza global.

Es indispensable una amplia participación de las partes interesadas, habida cuenta de las múltiples dimensiones éticas, jurídicas y técnicas de la IA. También hace falta que intervengan representantes de diversas disciplinas y culturas para garantizar que los sistemas de IA se ajusten al derecho internacional, el derecho humanitario, los derechos humanos y los compromisos de desarme antes de utilizarlos en el ámbito militar.

En cuarto lugar, es fundamental seguir teniendo presentes los riesgos, retos e implicaciones del desarrollo, despliegue y uso de la IA en el ámbito militar, sea o no de carácter tecnológico, y promover debates sustanciales al respecto. Indonesia destaca la importancia de evaluar continuamente las implicaciones generales de la IA militar para la paz y la seguridad internacionales, sobre todo en contextos de no proliferación y desarme. Se necesita investigar más a fondo para comprender estas repercusiones, que siguen estando insuficientemente estudiadas.

Determinar los riesgos asociados al desarrollo, despliegue y uso de la IA en el ámbito militar permitirá hacer pronósticos de base empírica, evaluar los riesgos y, dado el caso, concebir medidas para mitigarlos.

También es crucial mejorar la comprensión y el conocimiento de los riesgos asociados al uso de la IA en el ámbito militar. Debe promoverse la transparencia, entre otras cosas dando a conocer las políticas y estrategias nacionales, especialmente para determinar, evaluar y mitigar los riesgos; comunicando las capacidades de IA en el ámbito militar, cuando las haya, para potenciar la responsabilidad y fomentar la confianza; y compartiendo las lecciones aprendidas y las mejores prácticas a través de las fronteras, las industrias y los sectores.

En quinto lugar, la gobernanza de la IA no debe entorpecer el desarrollo tecnológico ni limitar el acceso de las naciones en desarrollo a esta tecnología. En los marcos no deben imponerse condicionalidades o barreras que restrinjan el acceso equitativo. Hace falta un enfoque equilibrado que aborde riesgos como la

proliferación y, al mismo tiempo, garantice la accesibilidad de la IA a los Estados con recursos limitados.

Por último, la gobernanza de la IA debe poner gran empeño en salvar la brecha digital y en materia de IA. Los países en desarrollo tienen grandes limitaciones, no solo en lo que se refiere a la IA, sino también en su capacidad de gobernar estas tecnologías de forma eficaz. Si no se subsana esta brecha, las iniciativas de gobernanza global se verán socavados, ya que muchos Estados siguen sin tener los medios necesarios para hacer frente a los retos complejos de carácter transfronterizo que plantea la IA.

Indonesia subraya la necesidad imperiosa de abordar las grandes brechas digitales y de IA entre las naciones y dentro de ellas, especialmente en lo que respecta al acceso a los recursos financieros, humanos y técnicos. Esas brechas podrían ahondar las desigualdades globales e incrementar los riesgos de conflicto.

La paz y la seguridad, que son bienes públicos globales, requieren la cooperación internacional de todos los países, desarrollados y en desarrollo, para abordar retos comunes y aprovechar beneficios colectivos, como los asociados al desarrollo, la aplicación y el uso de la IA en el ámbito militar. En este contexto, Indonesia pide una cooperación y una asistencia internacionales reforzadas y equilibradas para promover capacidades y marcos de gobernanza mundiales en materia de IA. Esa cooperación debe basarse en la equidad y la concertación y tener en cuenta las necesidades y situaciones específicas de los países en desarrollo, para lo que, entre otras medidas, habrá que emprender iniciativas de creación de capacidades, educación, transferencia de tecnología, aprendizaje permanente, capacitación técnica, investigación conjunta e intercambio de conocimientos.

También deberá establecerse en diversos niveles, no solo entre Estados y organizaciones internacionales, sino también entre sectores dentro de los países. Deben fomentarse las alianzas público-privadas para promover la innovación responsable y concienciar a la industria de las implicaciones que sus tecnologías pueden tener para la paz y la seguridad internacionales.

La cooperación internacional es fundamental no solo para cerrar las brechas digital y de IA, sino también para crear un entorno propicio para el fomento de la confianza entre Estados. Puede ayudar a reducir la división geopolítica y la competencia en materia de la IA. La cooperación internacional debe basarse en los principios de igualdad, confianza, mutuo beneficio, respeto de la soberanía y solidaridad a fin de allanar el camino a una colaboración sustancial, que comprenda la transferencia de tecnologías y el intercambio de conocimientos.

Indonesia también reconoce la utilidad de reforzar mecanismos de cooperación regional que tengan en cuenta las especificidades locales y regionales. Estos mecanismos pueden ser la base de un consenso global más amplio y aportar un espacio para deliberaciones más pormenorizadas y sensibles al contexto.

Irán (República Islámica del)

[Original: inglés]
[12 de marzo de 2025]

En respuesta a la solicitud de recabar las oportunidades y los retos que plantea para la paz y la seguridad internacionales la aplicación de la inteligencia artificial (IA) en el ámbito militar, con atención específica a los ámbitos distintos de los sistemas de armas autónomos letales, que la Asamblea General hizo al Secretario

General en el párrafo 7 de su resolución 79/239, la República Islámica del Irán presenta a continuación sus opiniones.

La inteligencia artificial, que se está convirtiendo en uno de los principales motores de cambio en el mundo actual, marcará de manera indeleble la evolución de la industria militar en el futuro cercano, lo que afectará a la esencia misma de la paz y la seguridad internacionales. Los actores estatales y no estatales promueven activamente agendas contrapuestas en materia de IA, que no pueden dejarse sin regular. Teniendo en cuenta el protagonismo de los actores no estatales y la necesidad de encontrar un equilibrio entre los procedimientos y las tendencias de regulación e innovación, es vital que la facultad de regular siga siendo prerrogativa soberana de los Estados Miembros.

Desde un punto de vista sustantivo, como ha ocurrido con otras tecnologías utilizadas en el ciberespacio y el espacio ultraterrestre, la República Islámica del Irán apoya la aplicación exclusivamente pacífica de la IA, consciente de que, en circunstancias adecuadas, las entidades militares también pueden beneficiarse pacíficamente de los dividendos de la IA.

Habida cuenta de los distintos niveles de desarrollo de las naciones, es de vital importancia garantizar que la brecha digital no se extienda a la IA. La inclusión de todos los procedimientos de regulación de la IA solo es posible en el marco de las Naciones Unidas, que se basa en el consenso. Este enfoque protege la soberanía de los Estados Miembros, fomenta unas condiciones de desarrollo de la IA equitativas para todos y aporta una flexibilidad en materia de innovación que permite prosperar al sector de la IA. La centralidad de las Naciones Unidas en las cuestiones de regulación de la IA impide planteamientos de exclusividad nacional en la materia. En este asunto de vital importancia deben imperar la inclusión y los planteamientos de consenso.

A pesar de los debates sobre la IA que se están manteniendo en diversos foros internacionales, nuestra comprensión de la cuestión y de sus implicaciones para la paz y la seguridad internacionales sigue siendo incompleta. Es prematuro afirmar que el derecho internacional, el derecho humanitario y el derecho internacional de los derechos humanos es plenamente aplicable a la IA. Ante un fenómeno nuevo y en rápida evolución de tal magnitud, el marco jurídico internacional podría necesitar una adaptación y una evolución singulares.

En lo que respecta a las iniciativas internacionales de regulación, la República Islámica del Irán es partidaria de establecer acuerdos jurídicamente vinculantes entre Estados Miembros, una línea de actuación que considera preferible a los instrumentos normativos o políticos.

Desde su posición de principios en materia de desarme, la República Islámica del Irán rechaza todo planteamiento políticamente motivado, discriminatorio o condicional, así como los dobles raseros. Por ello, los términos que utilice la Asamblea General han de reflejar un sentido de unidad y consenso. Conceptos como “aplicación responsable” son demasiado abstractos para regular un ámbito caracterizado por la concreción y la exactitud. Una noción tan abstracta se prestaría a malas interpretaciones y podría dar pie a planteamientos politizados. La República Islámica del Irán expresa su firme objeción al uso de un término tan subjetivo y preconiza y propone sustituir la expresión “aplicación responsable” por “aplicación pacífica” en cualquier instrumento futuro.

Israel

[Original: inglés]
[10 de abril de 2025]

Israel toma nota de la aprobación de la resolución [79/239](#) de la Asamblea General y, en respuesta a su párrafo 7, tiene el honor de presentar su contribución nacional al informe que el Secretario General presentará a la Asamblea en su octogésimo período de sesiones para que los Estados Miembros prosigan su examen.

Israel estima que el concepto de inteligencia artificial (IA) es objeto en la actualidad de diversas interpretaciones posibles que pueden irse matizando con el tiempo.

Es obvio que el uso de la IA en el ámbito militar es cada vez más común y frecuente. Israel votó a favor de la resolución de la Asamblea General antes señalada y alienta a los Estados y a todas las partes interesadas a entablar, con profesionalidad y sin politicización, un debate que tenga en cuenta las consideraciones legítimas de todos los Estados, incluidas las consideraciones de seguridad, humanitarias, económicas y de desarrollo.

Para mantener un debate serio y responsable sobre la IA en el ámbito militar que, en última instancia, pueda tener efectos sustanciales, creemos que debe adoptarse un enfoque pragmático, equilibrado y gradual.

Como la tecnología ofrece una amplia variedad de oportunidades en casi todos los ámbitos, incluido el militar, celebramos que se estudien los beneficios que pueden aportar estos avances y las formas de concretarlos, así como sus posibles riesgos y las formas de mitigarlos. Israel considera que las tecnologías emergentes, como las basadas en la IA, pueden servir también para promover el cumplimiento del derecho internacional humanitario vigente. Esas posibles oportunidades obligan a no estigmatizar negativamente esas tecnologías.

Israel, que sigue siendo una voz constructiva en el debate global sobre la aplicación militar de la IA, respaldó recientemente la Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar, promovida por los Estados Unidos. Esperamos sumarnos a futuras reuniones de esta iniciativa y seguir fomentando el uso militar responsable de la IA y la autonomía.

En la Declaración, así como en otros contextos, los Estados han considerado en los últimos años la orientación del desarrollo y el uso de la inteligencia artificial en el ámbito militar, ya sea a escala nacional o internacional. A continuación se indican algunos de los principios más básicos y comúnmente aceptados de tal orientación, que también pueden ser de interés para las deliberaciones que tengan lugar en el marco de la resolución [79/239](#):

- El uso militar de la IA debe ajustarse al derecho internacional vigente.
- Debe ser responsable y mejorar la seguridad internacional.
- Los Estados deben garantizar que se rinden cuentas por el uso de capacidades de IA conforme al derecho internacional vigente, por ejemplo utilizándolas en el marco de una cadena responsable de mando y control humanos.

A continuación se indican algunas de las prácticas que los Estados deben adoptar para hacer efectivos estos principios:

- Los Estados deben tomar las medidas adecuadas, por ejemplo realizar exámenes jurídicos, para garantizar que sus capacidades militares de IA se utilizarán con

arreglo a las obligaciones que les incumben en virtud del derecho internacional, en particular el derecho internacional humanitario.

- Los Estados deben tomar las medidas adecuadas para garantizar el desarrollo, despliegue y uso responsables de las capacidades militares de IA. Estas medidas deben aplicarse en las fases que corresponda del ciclo de vida de dichas capacidades.
- El personal competente debe actuar con la debida diligencia al desarrollar, desplegar y utilizar las capacidades militares de IA y los sistemas de armamentos que las incorporen.
- Los altos funcionarios deben supervisar de forma eficaz y adecuada el desarrollo y despliegue de las capacidades militares de IA cuya utilización pueda tener consecuencias importantes y los sistemas de armas que incorporen tales capacidades, entre otros.
- Los Estados deben promover las actuaciones adecuadas para garantizar que las capacidades militares de IA se utilicen de forma responsable y legal, y tratar constantemente con otros Estados la manera en que se despliegan y utilizan las capacidades militares de IA.

Israel considera provechoso mantener debates multilaterales inclusivos sobre la IA en el ámbito militar y sus implicaciones para la seguridad internacional que logren el equilibrio adecuado entre la necesidad militar y las consideraciones humanitarias.

Italia

[Original: inglés]
[11 de abril de 2025]

Presidencia italiana del Grupo de los Siete

En 2024, durante la presidencia italiana del Grupo de los Siete, la inteligencia artificial (IA) ocupó el centro de los debates políticos y técnicos. La cumbre de dirigentes celebrada en Apulia reconoció la impronta de la IA en el ámbito militar y la necesidad de un marco de desarrollo y uso responsables.

Del 18 al 20 de octubre de 2024 se celebró en Nápoles la primera reunión ministerial del Grupo de los Siete dedicada a la defensa. En esa ocasión, los Ministros de Defensa del Grupo de los Siete reafirmaron su determinación de abordar los retos de seguridad de forma cohesionada y concreta, en un momento histórico de gran inestabilidad. Además, subrayaron la necesidad de un enfoque más cooperativo de investigación y desarrollo en materia de defensa, que entraña compartir y aprovechar la experiencia y los conocimientos fomentando al mismo tiempo un entorno seguro que permita evitar accesos con malas intenciones a fin de mantener la ventaja competitiva en ámbitos como las tecnologías emergentes y disruptivas.

Por último, en su declaración, el Grupo de Directores de No Proliferación del Grupo de los Siete reconoció el profundo impacto de las tecnologías disruptivas emergentes, como la IA, en el control de armamentos, la no proliferación y el desarme, así como en el futuro de las operaciones militares.

I. Inteligencia Artificial Responsable en el Ámbito Militar

Italia valora el proceso de Inteligencia Artificial Responsable en el Ámbito Militar iniciado por los Países Bajos y la República de Corea en 2023 con el propósito de ofrecer un foro de debate sobre las principales oportunidades, retos y riesgos asociados a las aplicaciones militares de la IA. En la Segunda Cumbre sobre la

Inteligencia Artificial Responsable en el Ámbito Militar, celebrada en Seúl en 2024, Italia respaldó el Plan de Acción, un documento en el que se esbozan principios clave para una gobernanza responsable de la IA, como el cumplimiento del derecho internacional, la responsabilidad y la rendición de cuentas humanas, la fiabilidad y solvencia de los sistemas de IA y la implicación humana adecuada en el desarrollo, despliegue y uso de la IA en el ámbito militar.

Los Estados que han respaldado el Plan subrayan la necesidad de evitar que las tecnologías de IA se utilicen para contribuir a la proliferación de armas de destrucción masiva y que se socaven las labores de control de armamentos, desarme y no proliferación. Además, para llegar un entendimiento común respecto de la IA y sus aplicaciones en el ámbito militar, en el Plan se insta a los Estados a que se comprometan a seguir examinando la cuestión y a establecer procedimientos eficaces de examen jurídico, medidas de fomento de la confianza y medidas adecuadas de reducción de riesgos. En este marco, el intercambio de información y buenas prácticas, así como la participación activa de otras partes interesadas, son cruciales para hacer avanzar el debate.

II. Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar

Italia también valora la Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar. Afirmando que el uso militar de la IA podía y debía ser ético y responsable y mejorar la seguridad internacional, los Estados que la respaldaron reconocieron la necesidad de aplicar una serie de medidas respecto del desarrollo, despliegue y uso de las capacidades militares de IA. Concretamente, se comprometieron a adoptar las siguientes: minimizar los sesgos involuntarios en las capacidades militares de IA; garantizar que su seguridad y eficacia se sometieran a pruebas adecuadas y rigurosas; y establecer las debidas salvaguardias para poder detectar y evitar consecuencias indeseadas y actuar con eficacia cuando las haya. Además, es importante definir una cadena humana responsable de mando y control y utilizar las capacidades militares de IA con arreglo a las obligaciones internacionales.

III. El Pacto para el Futuro

En septiembre de 2024, los dirigentes mundiales aprobaron el Pacto para el Futuro, con lo que reafirmaron sus compromisos globales e hicieron posible que los Estados abordaran retos y oportunidades nuevos e incipientes. En la acción 27, se alienta a los Estados a aprovechar las oportunidades que ofrecen tecnologías nuevas y emergentes como la IA y a abordar los riesgos que puede plantear su uso indebido. Concretamente, los Estados Miembros seguirán evaluando esos riesgos y las posibles oportunidades de las aplicaciones militares de la IA durante todo su ciclo de vida, en consulta con las instancias pertinentes.

IV. Declaración de París sobre el mantenimiento del control humano en los sistemas de armas dotados de IA

Italia también respaldó recientemente la Declaración de París sobre el mantenimiento del control humano en los sistemas de armas dotados de IA, que se aprobó al margen de la Cumbre de Acción sobre Inteligencia Artificial celebrada en París del 6 al 11 de febrero de 2025. Tras subrayar que la responsabilidad y la rendición de cuentas nunca pueden delegarse en las máquinas, los Estados signatarios se comprometieron a adoptar un enfoque de desarrollo, despliegue y uso de aplicaciones de IA en el ámbito militar centrado en el ser humano. También se comprometieron a garantizar que el despliegue de la IA en el sector militar se ajuste

plenamente al derecho internacional y al derecho internacional humanitario y a promover la investigación, el desarrollo y la innovación con tecnologías de IA.

V. Grupo de Expertos Gubernamentales sobre las Tecnologías Emergentes en el Ámbito de los Sistemas de Armas Autónomos Letales

Los rápidos avances en inteligencia artificial y aprendizaje automático también tienen importantes consecuencias para la autonomía de los sistemas de armas. En opinión de Italia, la Convención sobre Prohibiciones o Restricciones del Empleo de Ciertas Armas Convencionales que Puedan Considerarse Excesivamente Nocivas o de Efectos Indiscriminados, que reúne los conocimientos diplomáticos, jurídicos y militares de representantes gubernamentales, organizaciones internacionales y organismos especializados, es, con mucho, el foro más indicado para tratar las cuestiones actuales y emergentes de desarrollo y utilización de sistemas de armas. Italia participa activamente en los debates del Grupo de Expertos Gubernamentales sobre las Tecnologías Emergentes en el Ámbito de los Sistemas de Armas Autónomos Letales, que se estableció bajo los auspicios de la Convención y está dedicado a hacer avanzar los debates sobre la elaboración de los componentes de un futuro instrumento, conforme al mandato aprobado por la Reunión de 2023 de las Altas Partes Contratantes en la Convención.

A juicio de Italia, ese instrumento debería establecer prohibiciones y normativas claras para que, llegado el momento, pudiera aprobarse como Protocolo Adicional de la Convención sobre Ciertas Armas Convencionales. Según este enfoque, los sistemas de armas autónomos letales que no puedan desarrollarse y utilizarse de conformidad con el derecho internacional humanitario estarían prohibidos *ipso facto*. Por otra parte, se regularían los sistemas dotados de autonomía de decisión en las funciones críticas, que pueden desarrollarse y utilizarse respetando plenamente el derecho internacional humanitario. En opinión de Italia, el elemento humano es de hecho crucial para todo el ciclo de vida de los sistemas de armas autónomas letales, es decir, para su concepción, desarrollo, producción, despliegue y uso. También ha de mantenerse un nivel adecuado de juicio y control humanos a fin de garantizar la responsabilidad y la rendición de cuentas con arreglo al derecho internacional humanitario.

Japón

[Original: inglés]
[11 de abril de 2025]

En su resolución 79/239, la Asamblea General solicitó al Secretario General que recabase las opiniones de los Estados Miembros y de los Estados observadores sobre las oportunidades y los retos que planteaba para la paz y la seguridad internacionales la aplicación de la inteligencia artificial (IA) en el ámbito militar, con atención específica a los ámbitos distintos de los sistemas de armas autónomos letales, y que, en su octogésimo período de sesiones, le presentase un informe sustantivo en el que se resumieran esas opiniones y se clasificaran las propuestas normativas existentes y emergentes, con un anexo en el que figurasen esas opiniones, para que los Estados siguieran debatiendo. El Japón presenta a continuación sus opiniones al respecto con el propósito de contribuir a la elaboración del informe y al desarrollo de los debates sobre el tema.

I. Opiniones generales

El Japón tiene el compromiso de mantener y reforzar un orden internacional libre y abierto fundado en el estado de derecho para que todos los pueblos puedan gozar de paz, estabilidad y prosperidad y de promover la diplomacia para hacer realidad un mundo seguro en el que se proteja la dignidad humana. En consonancia con estos objetivos, el Japón ha participado activamente en iniciativas de fomento de la paz y la seguridad internacionales y de control de armamentos y desarme.

El Japón opina que la aplicación de la IA en el ámbito militar debe examinarse globalmente, con una comprensión suficiente de sus riesgos y beneficios y teniendo en cuenta las consideraciones humanitarias y las perspectivas de seguridad. Es útil profundizar en la comprensión de la aplicación de la IA en el ámbito militar y promover medidas realistas y prácticas para su uso responsable a fin de aprovechar al máximo sus beneficios y reducir los riesgos.

Respecto de la aplicación de la IA en el ámbito militar, el Japón comparte la opinión de que, en primer lugar, el derecho internacional vigente se aplica a los asuntos de su competencia que tengan lugar durante todo del ciclo de vida de la IA; en segundo lugar, las capacidades de IA deben utilizarse de manera responsable; y, en tercer lugar, los seres humanos siguen teniendo la responsabilidad y la obligación de rendir cuentas sobre su uso y efectos. El Japón también hace hincapié en la necesidad de mejorar la transparencia, que es una importante medida de fomento de la confianza, para aprovechar al máximo los beneficios y reducir los riesgos.

II. Opiniones y visión del Japón sobre las oportunidades y los retos para la paz y la seguridad internacionales de la aplicación de la inteligencia artificial en el ámbito militar

Oportunidades

Opiniones

Los rápidos avances de la ciencia y la tecnología, incluida la IA, están transformando radicalmente el paradigma de la seguridad. Los países se esfuerzan por desarrollar tecnologías de vanguardia que podrían alterar drásticamente la naturaleza de la guerra y acabar siendo determinantes y, en la práctica, se ha hecho extremadamente difícil distinguir entre las tecnologías para uso civil y las destinadas a fines de seguridad. La IA encierra un potencial extraordinario para transformar todos los aspectos de los asuntos militares, incluidas las operaciones militares; las actividades de mando y control; las actividades de inteligencia, vigilancia y reconocimiento; el adiestramiento; la gestión de la información; y el apoyo logístico. Habida cuenta de los diversos usos de la IA en el ámbito militar, con su aplicación puede conseguirse, entre otros beneficios, aumentar la precisión, la exactitud y la eficacia; mejorar la conciencia y la comprensión situacionales; analizar información con mayor rapidez; reducir los errores humanos; y ahorrar mano de obra. Una aplicación adecuada podría contribuir a una mejor protección de los civiles en los conflictos y a consolidar la paz cuando terminan.

Modo en que el Japón aborda el aprovechamiento de las “oportunidades”

Al aplicar la IA en el ámbito militar, hay que plantearse si es una forma eficaz de superar problemas observados por los humanos sin perder de vista las funciones y limitaciones de la IA. Aplicar la IA no es de por sí un objetivo y no ha de plantearse sin considerar las funciones y limitaciones de esta tecnología. Por tanto, los Estados deben asegurarse de que las capacidades militares de IA tengan usos explícitos y bien

definidos y de que estén concebidas y desarrolladas para cumplir las funciones previstas. Teniendo presente esta cuestión, es importante promover una comprensión internacional común de la IA y de sus funciones y limitaciones en el ámbito militar, y una comprensión común de la posible aplicación de la IA en el ámbito militar. En cuanto a la aplicación de la IA por parte de las autoridades de defensa, el Ministerio de Defensa del Japón publicó en julio de 2024 la Política Básica del Ministerio de Defensa sobre la Promoción de la Utilización de la IA, en la que expone sus ideas actuales sobre las funciones y limitaciones de la IA en el ámbito militar y los ámbitos de aplicación de esta tecnología que considera prioritarios. A la luz de las capacidades y limitaciones actuales de la IA, el Ministerio de Defensa define en ese documento los siete ámbitos en los que centra la aplicación de la IA, a saber:

- La detección e identificación de objetivos
- La recopilación y el análisis de información de inteligencia
- Las tareas de mando y control
- Las operaciones de apoyo logístico
- Los sistemas no tripulados
- La ciberseguridad
- Una labor administrativa más eficiente

En la Política Básica también se señala que conviene tener presente que la IA se utiliza para apoyar la toma de decisiones humanas y que la participación humana es esencial al aplicarla.

Desafíos

Opiniones

La aplicación de la IA en el ámbito militar puede entrañar riesgos de uso indebido o malintencionado y de intensificación y reducción del umbral de los conflictos, lo que pueden deberse a sesgos, consecuencias indeseadas y otros factores. A este respecto, el Japón subraya la necesidad de evitar que haya Estados y agentes no estatales que utilicen la IA para contribuir a la proliferación de armas de destrucción masiva y hace hincapié en que la IA no debe entorpecer, sino apoyar, las iniciativas de desarme, control de armamentos, y no proliferación.

Modo en que el Japón aborda el tratamiento de los “desafíos”

En cuanto a los riesgos como los sesgos y los usos indebidos o malintencionados que plantea la IA, el Ministerio de Defensa del Japón tratará de reducirlos tomando como referencia los conceptos de IA centrada en el ser humano, seguridad, equidad, protección de la privacidad, garantía de seguridad, transparencia y rendición, que se establecen en las Directrices sobre IA para empresas en el Japón publicadas en abril de 2024 y tendrá en cuenta los debates que mantenga en la comunidad internacional y con autoridades de defensa de otros países.

Además, el Japón sigue de cerca el posible efecto de las tecnologías emergentes como la IA en el desarme nuclear y la no proliferación. A este respecto, el Japón acoge con satisfacción el compromiso contraído por los Estados Unidos, el Reino Unido y Francia en la Conferencia de las Partes de 2022 encargada del Examen del Tratado sobre la No Proliferación de las Armas Nucleares de mantener el control y la implicación humanos en todas las acciones críticas para informar y ejecutar decisiones soberanas relativas al empleo de armas nucleares e insta a otros Estados poseedores de armas nucleares a seguir el ejemplo. Además, en su recomendación a

la Conferencia de Examen de 2026, el Grupo Internacional de Personas Eminentas por un Mundo sin Armas Nucleares subrayó la necesidad de colaborar en el tratamiento de los retos y oportunidades asociados a las tecnologías emergentes.

III. Opiniones sobre el futuro de los debates y la cooperación internacional

Hace falta un planteamiento flexible, equilibrado y realista de gobernanza de la IA en el ámbito militar para seguir el rápido ritmo de desarrollo y avance de las tecnologías. El Japón subraya que las iniciativas para una IA responsable en el ámbito militar pueden realizarse en paralelo a las labores de investigación, desarrollo, experimentación e innovación en esta tecnología, sin entorpecerlas.

Cabe señalar que los debates que problematizan tecnologías específicas de IA pueden obstaculizar el desarrollo y la innovación tecnológicos en el sector civil y llegar a tener un efecto paralizador. Además, la aplicación de la IA en el ámbito militar debe abordarse de forma inclusiva que entrañe la participación e interacción de los interesados.

A la luz de estas consideraciones, el Japón apoya firmemente los resultados de las Cumbres sobre la Inteligencia Artificial Responsable en el Ámbito Militar y la Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar y espera que más Estados se sumen a esas iniciativas.

En cuanto a los sistemas de armas autónomos letales, cabe señalar que el Japón apoya firmemente la continuación de los debates en el marco de la Convención sobre Prohibiciones o Restricciones del Empleo de Ciertas Armas Convencionales que Puedan Considerarse Excesivamente Nocivas o de Efectos Indiscriminados y espera que los debates sobre la aplicación de la IA en el ámbito militar complementen y refuerzen los del Grupo de Expertos Gubernamentales sobre las Tecnologías Emergentes en el Ámbito de los Sistemas de Armas Autónomos Letales establecido en el marco de la Convención.

También reconoce que la transparencia en la aplicación de la IA en el ámbito militar es una importante medida de fomento de la confianza que permite reducir riesgos, al igual que la colaboración y cooperación efectivas entre países. Reconoce asimismo que es importante crear capacidades para facilitar un enfoque responsable de desarrollo, despliegue y uso de la IA en el ámbito militar y se compromete a reforzar la cooperación internacional en materia de creación de capacidades con el fin de reducir la brecha de conocimientos en relación con ese enfoque. A este respecto, serán útiles métodos como el intercambio de buenas prácticas y lecciones aprendidas y el Japón aprovechará las posibilidades de intercambiar opiniones con otros países.

Por último, en lo que respecta a la aplicación de la IA en el ámbito militar, el Japón seguirá participando de manera activa y constructiva en los debates internacionales con el objetivo de que la comunidad internacional llegue a un entendimiento común mediante discusiones equilibradas que tengan en cuenta las consideraciones humanitarias y las perspectivas de seguridad.

Lituania

[Original: inglés]
[9 de abril de 2025]

Lituania agradece la oportunidad de presentar una comunicación para el informe del Secretario General en respuesta a la resolución [79/239](#), que la Asamblea General aprobó el 24 de diciembre de 2024 y que Lituania se complació en apoyar.

Lituania observa que el desarrollo y el uso de la inteligencia artificial (IA) en el ámbito militar entraña tanto oportunidades como desafíos para la paz y la seguridad internacionales. Lituania concede gran importancia a la elaboración de normas y principios de uso responsable que permitan a los Estados aprovechar los beneficios y mitigar posibles riesgos de la IA en el ámbito militar y tiene la firme convicción de que todos los Estados responsables debería garantizar la aplicación responsable de la IA en el ámbito militar por su propio interés y de que, para abordar las implicaciones de la IA militar, se precisa una acción global y un enfoque de múltiples partes interesadas que de cabida a los sectores público y privado, a la sociedad civil y al mundo académico.

Lituania apoya firmemente la Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar, a la que se sumó el 13 de noviembre de 2023. En ella se formulan principios jurídicamente no vinculantes y mejores prácticas para garantizar un uso responsable y legal de la IA en el ámbito militar. Contempla medidas como realizar exámenes jurídicos, hacer la supervisión adecuada, reducir al mínimo los sesgos involuntarios y garantizar que las capacidades militares de IA tengan uso explícitos y bien definidos. Lituania alienta encarecidamente a otros Estados a que la suscriban.

Además, Lituania suscribe la estrategia de IA de la Organización del Tratado del Atlántico Norte, aprobada en 2021 y revisada en 2024, en la que se establecen seis principios de uso responsable de la IA en el ámbito militar, a saber: legalidad, responsabilidad y rendición de cuentas, explicabilidad y trazabilidad, fiabilidad, gobernabilidad y mitigación de sesgos. Estos principios jurídicamente no vinculantes, a los que Lituania se ha adherido, han de aplicarse durante todo el ciclo de vida de las aplicaciones de IA.

Por último, Lituania se complace en presentar su opinión sobre las oportunidades y los retos para la seguridad internacional que plantea la aplicación la IA en el ámbito militar. A su juicio, la IA militar podría y debería utilizarse de forma responsable, ante todo, para reforzar la seguridad nacional del Estado y contribuir a la aplicación del derecho internacional, incluido el derecho internacional humanitario, y al cumplimiento de las diversas obligaciones del Estado relativas a la protección de los civiles. Además de reforzar la protección de los civiles en los conflictos armados, la IA responsable ofrece oportunidades para mejorar la toma de decisiones, la logística, la planificación y otras operaciones que aumentan la eficacia.

Por lo que respecta a los posibles riesgos de la IA en el ámbito militar, Lituania destacaría los relativos a la ciberseguridad, los sesgos involuntarios en las capacidades militares de IA y los comportamientos imprevistos de los sistemas que funcionan con esta tecnología. Lituania cree que la mejor manera de abordarlos es aplicar principios de uso responsable, crear capacidades y capacitar adecuadamente al personal en el uso de las aplicaciones de IA y los sistemas que funcionan con IA. Lituania hace hincapié en que, para sacar partido de la IA en el ámbito militar y utilizarla como una capacidad crítica de defensa, los Estados no deben imponer restricciones innecesarias excesivas que entorpezcan la innovación en materia de IA, sobre todo si hay Estados irresponsables que se niegan a aceptar tales restricciones.

México

[Original: español]
[10 de abril de 2025]

Artificial intelligence, autonomous weapons systems and the challenge facing the world in terms of their regulation

México presenta este documento en cumplimiento de la resolución [79/239](#) de la Asamblea General, intitulada “La inteligencia artificial en el ámbito militar y sus implicaciones para la paz y la seguridad internacionales”.

México reconoce que la aplicación de la inteligencia artificial (IA) en el ámbito militar puede ofrecer beneficios. Sin embargo, también plantea desafíos significativos para la paz y la seguridad internacionales, que requieren atención urgente y coordinada por parte de la comunidad internacional.

México valora los intercambios multilaterales impulsados en el marco de las Naciones Unidas, como el “Diálogo inaugural sobre IA militar: Oportunidades, riesgos y paz y seguridad internacionales”, coordinado por la Oficina de Asuntos de Desarme de las Naciones Unidas, que contribuyen a generar un entendimiento común sobre los riesgos emergentes y las responsabilidades compartidas. Coincidimos en que la integración de la IA en aplicaciones militares plantea desafíos fundamentales para la paz y la seguridad internacionales, incluida la escalada involuntaria de conflictos, la ambigüedad estratégica y la creciente autonomía en el uso de la fuerza.

México considera prioritario fortalecer la cooperación internacional, fomentar la transparencia, compartir buenas prácticas, crear capacidades hacia una cultura de cumplimiento normativo y de respeto al derecho internacional, así como avanzar en la formulación de marcos normativos que aseguren que el desarrollo y despliegue de la IA en contextos militares se rija por principios éticos, jurídicos y humanitarios, evitando que esta tecnología profundice asimetrías o erosione la estabilidad internacional.

Paz y seguridad internacionales

Para México es imperativo tomar acciones para evitar la proliferación y el uso indebido de estas tecnologías, inclusive por actores no estatales o fuera de marcos jurídicos claros.

La paz y la seguridad internacionales no pueden supeditarse a la incorporación de las tecnologías nuevas y emergentes en el ámbito militar. Dicha incorporación debe responder a una perspectiva de desarrollo humano y de empoderamiento social, de manera particular en favor de los países en desarrollo. Por tanto, el objetivo de este tipo de tecnologías debe orientarse a sus usos pacíficos y de solución de controversias, en lugar de buscar que la maquinaria militar sea más eficiente.

La creciente sofisticación de las amenazas digitales y la posibilidad de que las tecnologías emergentes sean empleadas como vectores de ataque de Estado a Estado, así como la dificultad para garantizar la fiabilidad y precisión de sistemas autónomos en contextos militares, la exposición a vulnerabilidades en el ciclo de vida de la IA, los sesgos algorítmicos, el envenenamiento de datos y el uso de modelos generativos para fines maliciosos, subrayan la necesidad de una mitigación proactiva de los riesgos.

Los avances científicos y tecnológicos, especialmente en IA, sistemas autónomos, tecnologías digitales y cuánticas, superan la capacidad actual de los marcos normativos para gestionar dichos riesgos. Por ello, México reitera la

necesidad de desarrollar marcos de gobernanza integrales, fomentar la cooperación internacional y el diálogo multilateral y priorizar la transparencia, la rendición de cuentas y el control humano significativo a lo largo de todo el ciclo de vida de estas tecnologías, incluyendo pruebas rigurosas y salvaguardias éticas para su despliegue.

En ausencia de marcos jurídicos internacionales claros y del necesario consenso multilateral, el uso del término “responsable” en este contexto no debe interpretarse como una validación o aceptación tácita del uso o desarrollo de capacidades militares autónomas habilitadas por la IA. El principio de responsabilidad debe estar necesariamente vinculado a la legalidad y a la rendición de cuentas.

Bajo esa lógica, para México es esencial establecer mecanismos de gobernanza y regulación que reduzcan la probabilidad de que la IA y otras tecnologías disruptivas sean utilizadas para fines hostiles, reconociendo que los riesgos no se limitan únicamente a su despliegue operativo, sino que emergen desde las etapas iniciales de su diseño y desarrollo.

Contextos operacionales

Tomando en consideración los distintos contextos operacionales del ámbito militar en los cuales se podría incorporar esta tecnología, México observa que la IA puede tener impactos diferenciados.

En el contexto de los conflictos armados, resulta imperativo asegurar que toda tecnología basada en la IA se utilice en conformidad con el derecho internacional humanitario, en particular con los principios de distinción, proporcionalidad, precaución y humanidad.

En el ámbito de las operaciones de mantenimiento de la paz y respuesta a desastres, la IA puede contribuir positivamente a la coordinación logística, la predicción de riesgos y la atención a poblaciones afectadas, siempre que se respete plenamente el marco de los derechos humanos.

En lo que respecta a la seguridad fronteriza, México reconoce que la IA puede fortalecer las capacidades de monitoreo; sin embargo, subraya la importancia de garantizar el respeto a la dignidad de todas las personas, evitando decisiones automáticas que perpetúen prácticas discriminatorias.

Sistemas de armas autónomos letales

México considera que una parte fundamental de esta discusión son los sistemas de armas autónomos letales, los cuales plantean una preocupación particular para la paz y seguridad internacionales. En ese sentido, abogamos por no fragmentar las discusiones multilaterales sobre la incorporación de nuevas tecnologías en el ámbito militar y consideramos que los sistemas de armas autónomos letales deben ser parte integral de estos intercambios.

México considera urgente que la comunidad internacional establezca prohibiciones y regulaciones claras a los sistemas de armas autónomos letales, por su incompatibilidad con el derecho internacional humanitario y sus riesgos éticos, jurídicos y de seguridad.

México ha impulsado y copatrocinado las resoluciones [78/241](#) y [79/62](#) sobre sistemas de armas autónomos letales en la Asamblea General, a fin de consolidar un espacio multilateral legítimo para atender estos desafíos.

México respalda el llamado del Secretario General y del Comité Internacional de la Cruz Roja para iniciar negociaciones de un instrumento jurídicamente vinculante

que establezca las prohibiciones y regulaciones necesarias sobre sistemas de armas autónomos letales hacia 2026, como se refleja en la Nueva Agenda de Paz.

México ha refrendado su compromiso político en este tema mediante su participación en la Conferencia de San José (2023), la adhesión al Comunicado de Belém, y su involucramiento activo en la conferencia “Humanity at the crossroads: autonomous weapons systems and the challenge of regulation” (La humanidad en la encrucijada: los sistemas de armas autónomos y el desafío de la regulación) (Viena, 2024), cuyo informe final fue endosado por nuestro país.

México estima que los sistemas de armas autónomos letales plantean múltiples riesgos, entre ellos:

- Excluyen el juicio humano en decisiones críticas sobre el uso de la fuerza
- Sustituyen la valoración contextual indispensable en operaciones militares
- Debilitan los mecanismos de rendición de cuentas y atribución de responsabilidad

La responsabilidad por el uso de la fuerza nunca debe ser transferida a una máquina. Las decisiones sobre despliegue, activación o anulación de sistemas armados deben recaer en todo momento en personas humanas, sujetas a responsabilidad legal.

México reitera que toda tecnología militar, incluida aquella basada en la IA, debe respetar las obligaciones internacionales derivadas de:

- La Carta de las Naciones Unidas
- El derecho internacional humanitario
- El derecho internacional de los derechos humanos
- El derecho penal internacional
- El derecho de la responsabilidad internacional

En ese marco, México considera indispensable prohibir aquellos sistemas de armas cuya tecnología:

- No pueda distinguir entre objetivos militares y civiles
- No pueda aplicar el principio de proporcionalidad en relación con los daños colaterales
- No cuente con mecanismos de cancelación si el ataque resulta improcedente
- Producen sufrimientos innecesarios o lesiones superfluas

México insiste en la urgencia de iniciar negociaciones de un instrumento jurídicamente vinculante, que establezca prohibiciones y regulaciones específicas sobre los sistemas de armas autónomos letales; garantice el mantenimiento del control humano significativo en actividades críticas; e incluya mecanismos de implementación, monitoreo y rendición de cuentas eficaces.

Beneficios y riesgos

Sobre los usos específicos de la IA, México identifica tanto beneficios como riesgos en los siguientes ámbitos:

- Mando y control: bajo ciertas condiciones, la IA podría mejorar la eficiencia de las decisiones operativas, pero estas deben permanecer bajo control humano significativo, especialmente cuando se trate del uso de la fuerza. Se reconoce

que la IA tiene la capacidad de procesar y analizar grandes volúmenes de datos e información, que superan con creces las capacidades humanas, lo que permite agilizar, facilitar y eficientizar la predicción de tendencias futuras y fundamentar decisiones estratégicas en tiempo real.

- Ciberoperaciones: la IA ofrece capacidades valiosas para la predicción y respuesta a incidentes cibernéticos, aunque también incrementa los riesgos de una escalada en las tensiones, incluyendo el uso ofensivo automatizado sin supervisión adecuada.
- Gestión de información y logística: el procesamiento masivo de datos mediante la IA puede facilitar decisiones en tiempo real, pero debe ir acompañado de protocolos que aseguren su uso ético, explicable y responsable.

No obstante lo anterior, México subraya los riesgos tecnológicos asociados a la integración de la IA en contextos militares, considerando que la evidencia sugiere la persistencia de fallas técnicas o errores no previstos que pueden escalar un conflicto.

Noruega

[Original: inglés]
[11 de abril de 2025]

Noruega acoge positivamente la posibilidad de presentar sus opiniones sobre las oportunidades y retos para la paz y la seguridad internacionales que entraña la aplicación de la inteligencia artificial (IA) en el ámbito militar, específicamente en sectores distintos de los sistemas de armas autónomas letales, conforme a lo dispuesto en la resolución [79/239](#) de la Asamblea General, titulada “La inteligencia artificial en el ámbito militar y sus implicaciones para la paz y la seguridad internacionales”.

Como se reconoce en el informe de políticas del Secretario General sobre la Nueva Agenda de Paz que se publicó en julio de 2023, la inteligencia artificial es una tecnología a la vez facilitadora y perturbadora de uso cada vez más común en una amplia gama de aplicaciones civiles, militares y de doble uso. La creciente ubicuidad de la inteligencia artificial, sumada a la rápida escalabilidad, la falta de transparencia y el ritmo de innovación que la caracterizan, puede entrañar riesgos para la paz y la seguridad internacionales y plantea problemas de gobernanza.

Noruega, firme defensora del derecho internacional, el multilateralismo y la innovación responsable en el sector de la defensa, apoya las iniciativas encaminadas a promover un entendimiento común, reforzar la gobernanza y regular adecuadamente la IA en el ámbito militar. Se parte del principio básico de que las aplicaciones de la IA en el ámbito militar han de desarrollarse, desplegarse y emplearse de forma responsable durante todo su ciclo de vida y atenerse al derecho internacional vigente, en particular al derecho internacional humanitario.

Cabe destacar que, en su resolución [79/239](#), la Asamblea General afirma que el derecho internacional, incluidos la Carta de las Naciones Unidas y el derecho internacional humanitario y de los derechos humanos, se aplica al uso de la inteligencia artificial en el ámbito militar, destacando la importancia de que tal uso sea responsable y esté centrado en el ser humano.

Como tecnología instrumental, la IA tiene un extraordinario potencial de transformación de los asuntos militares en todos sus aspectos, incluidos la adquisición, los equipos, los programas informáticos, las operaciones, las funciones de mando y control, las comunicaciones estratégicas, la vigilancia, la inteligencia, el adiestramiento, la gestión de la información y el apoyo logístico. La aplicación de la IA en el ámbito militar conlleva oportunidades y riesgos previsibles e imprevisibles

en los planos táctico y estratégico. Como tecnología de uso general, la IA es un multiplicador de fuerzas capaz de redefinir la manera de hacer la guerra. La convergencia tecnológica entre inteligencia artificial, neurotecnología, biología sintética e informática cuántica hace que la cuestión sea aún más compleja.

Es primordial que la IA se desarrolle, despliegue, utilice y gestione de forma responsable, conforme a principios éticos básicos, en estricto cumplimiento de las obligaciones a las que están sujetos los Estados en virtud del derecho internacional, incluido el derecho internacional humanitario y de los derechos humanos, y asumiendo como fundamental el cometido de detectar y mitigar los riesgos.

En la Estrategia Noruega de Inteligencia Artificial en el Sector de la Defensa (2023) se señalan importantes ámbitos, más allá de los sistemas de armas autónomos letales, en los que la IA puede hacer aportaciones constructivas, a saber:

- **Mejora de la conciencia situacional y el apoyo para la toma de decisiones.** Utilizar la IA es a la vez una posibilidad y una necesidad en los ámbitos de la inteligencia, la vigilancia y el reconocimiento, ya que no es posible analizar manualmente los ingentes volúmenes de datos, que no dejan de aumentar. La IA puede utilizarse para el cribado de los datos pertinentes, por ejemplo mediante labores de pretratamiento de datos, traducción automática o detección de objetos especiales en imágenes a fin de descubrir anomalías y repeticiones, o contrastando datos para detectar intentos de desinformación. Las mejoras en este ámbito podrían redundar en unas operaciones más eficaces y precisas y reducir la pérdida de vidas humanas.
- **Ciberdefensa.** Los beneficios de la digitalización y la dependencia cada vez mayor de las tecnologías de la información y las comunicaciones traen consigo nuevas vulnerabilidades. El espacio digital brinda a los agentes generadores de amenazas la posibilidad de filtrar datos, participar en actividades de espionaje y sabotaje y realizar campañas de influencia. La IA puede reforzar la capacidad del sector de la defensa de detectar, vigilar, gestionar y contrarrestar las amenazas digitales y de informar al respecto. Su uso permite, entre otras cosas, obtener con mayor rapidez una visión más completa de objetivos y relaciones complejos, recabar información de fuentes pertinentes y racionalizar el uso de análisis. Para que en el futuro sea posible detectar y evitar ataques digitales es indispensable ampliar el conocimiento y la experiencia sobre las amenazas digitales a que puede dar lugar la IA, por lo que esta ha de ser un aspecto central de las estrategias que desarrolle el sector para defenderse de esas amenazas con instrumentos actuales y futuros.
- **Logística.** El éxito y la eficacia de las operaciones militares dependen de un apoyo logístico eficaz. Optimizando la logística con sistemas que empleen la IA podrían mejorarse la capacidad operacional y el estado de preparación. Las aplicaciones de IA en el sector de la logística civil ya han progresado mucho. Es probable que muchas de ellas puedan adaptarse con facilidad para utilizarlas en el sector militar.
- **Actividades de apoyo.** Probablemente, muchas actividades de apoyo militar podrían mejorar y ganar en eficiencia utilizando inteligencia artificial, por ejemplo, tareas de apoyo y refuerzo de la capacidad operacional como las relativas al uso y mantenimiento del material, la adquisición, gestión y eliminación de materiales y edificaciones, la contratación, adiestramiento y gestión del personal y la prestación de servicios comunes, como los de contabilidad y archivo. La IA permitiría reforzar las actividades de apoyo gracias a un mejor aprovechamiento de los datos utilizados para realizar análisis y fundamentar decisiones, a la automatización de tareas y a la mayor capacidad

de manejar información y conocimientos, lo que haría posible evolucionar a un modelo de mantenimiento predictivo, mejorar la circulación de la información, introducir nuevos y mejores sistemas de apoyo a la gestión de los recursos humanos y optimizar la modelización de la evolución de los costos de materiales y edificaciones. Por tanto, introduciendo adecuadamente esta tecnología en las actividades de apoyo podría ahorrarse tiempo y ganarse en eficacia.

Además, la utilización de la inteligencia artificial en el ámbito militar puede mejorar la aplicación del derecho internacional humanitario y contribuir a la protección de los civiles y los bienes de carácter civil en situaciones de conflicto armado. También puede ser útil en actividades de consolidación y mantenimiento de la paz y mejorar las capacidades de verificación y supervisión para el control de armas, el desarme y otros regímenes de cumplimiento.

Aplicar esta tecnología al ámbito militar plantea asimismo retos inéditos, ya que sus vulnerabilidades intrínsecas pueden tener consecuencias fortuitas que socaven el control, la responsabilidad y la rendición de cuentas humanos efectivos. Al utilizar el aprendizaje profundo, los modelos de IA pueden ser difíciles de entender, explicar y predecir, lo que, entre otras cosas, podría volver más confusa e impredecible la dinámica de intensificación de los conflictos.

Han de establecerse las salvaguardias necesarias para que el desarrollo, el despliegue y el uso de la inteligencia artificial sigan sujetas a un control y una supervisión sustancialmente humanos, algo tanto más importante cuanto más se acerque la aplicación de esta tecnología a las operaciones de combate y uso de la fuerza, como en el caso de los sistemas de apoyo para la toma de decisiones. La rendición de cuentas y la responsabilidad sobre el uso y los efectos de la IA militar han de permanecer siempre en manos humanas.

Los sistemas de inteligencia artificial pueden ser muy sensibles a la calidad y la representatividad de los datos utilizados para fines de entrenamiento. Los posibles sesgos, la desinformación e información errónea o unos datos de entrenamiento incompletos pueden hacer que los modelos generen resultados inexactos o discriminatorios. El sesgo de automatización puede llevar a los usuarios humanos a confiar en exceso en los resultados del sistema.

Unas capacidades de respuesta altamente automatizadas o autónomas en el ámbito cibernetico —sobre todo cuando no integran los mecanismos humanos adecuados— pueden generar respuestas indeseadas y llevar a una rápida escalada.

Una mayor dependencia de la cibertecnología para tareas que antes se realizaban manualmente o con un grado elemental de automatización también entraña el riesgo de que sus vulnerabilidades se aprovechen con intenciones maliciosas. El recurso cada vez mayor a sistemas comerciales suscita preocupación por la dependencia de proveedores externos, la pérdida de control sobre las actualizaciones y otras vulnerabilidades inherentes a los sistemas privativos.

Los anteriores son meros ejemplos de posibles riesgos asociados a la aplicación de la inteligencia artificial en el ámbito militar, pero hay otras muchas incógnitas. En un contexto militar, estos factores pueden, juntos o por separado, socavar los resultados de una misión y plantear graves problemas jurídicos, éticos, humanitarios y militares.

En la Estrategia Noruega de Inteligencia Artificial en el Sector de la Defensa (2023) también se presentan los siguientes principios clave para el desarrollo y uso responsables de la inteligencia artificial:

- **Legalidad.** Las aplicaciones de inteligencia artificial deben concebirse y utilizarse de conformidad con el derecho internacional, incluido el derecho

internacional humanitario y de los derechos humanos. Todo Estado que estudie, desarrolle, adquiera o adopte una nueva arma o nuevos medios o métodos de guerra dependientes de la inteligencia artificial, deberá determinar si el derecho internacional de los derechos humanos o cualquier otra norma del derecho internacional al que tal Estado esté sujeto prohíbe su empleo en algunas circunstancias o por completo.

- **Responsabilidad y rendición de cuentas.** Ha de garantizarse que la responsabilidad y la obligación de rendir cuentas por el uso de la inteligencia artificial sean exclusivamente humanas. Debe determinarse con precisión quién tiene la facultad de tomar decisiones sobre el uso de sistemas de IA y quién es responsable de su uso efectivo.
- **Explicabilidad, inteligibilidad, trazabilidad.** Las aplicaciones de la IA deben ser suficientemente explicables, inteligibles, transparentes y trazables.
- **Capacitación.** Los operadores de IA deben tener la formación necesaria para entender cómo se comportan sus aplicaciones y cómo detectar comportamientos anómalos.
- **Fiabilidad y seguridad.** Las aplicaciones de la inteligencia artificial deben tener unos ámbitos de uso explícitos y bien definidos, en los que habrá de ponerse a prueba y someterse a controles la resiliencia, fiabilidad y seguridad de esas aplicaciones a lo largo de todo su ciclo de vida. Deben tener asimismo unos niveles adecuados de seguridad y estar protegidas contra las amenazas digitales.
- **Control.** Debe garantizarse el control humano adecuado. Los sistemas de inteligencia artificial deben contar con una interfaz de interacción humano-máquina adecuada a su uso previsto, que permita detectar y mitigar consecuencias indeseadas, y con medios para adoptar las medidas correctivas necesarias cuando el sistema no funcione según lo previsto.

La comunidad internacional tiene que abordar con mayor profundidad las aplicaciones militares de la inteligencia artificial y sus implicaciones para la paz y la seguridad, así como las medidas necesarias para un uso responsable de esta tecnología en el ámbito militar. Debe prestarse especial atención a los sistemas de apoyo a las operaciones de combate, como los que aplican inteligencia artificial a la conciencia situacional y el apoyo a la toma de decisiones, en los que los resultados y comportamientos no deseados y la pérdida del control humano adecuado pueden tener consecuencias particularmente lesivas. También hay que abordar la IA en modalidades de guerra híbrida, como las ciberoperaciones, la guerra electrónica y las operaciones de información, entre otras.

Noruega está decidida a reforzar la cooperación internacional en materia de intercambio de información y creación de capacidad. Al crear una base de conocimientos compartida, los Estados promoverían un entendimiento común, cerrarían brechas, aumentarían la transparencia y fomentarían la confianza. Para ello, Noruega alentaría la elaboración y publicación de estrategias nacionales y documentos políticos relacionados con las aplicaciones militares de la IA. Debe prestarse atención a las medidas de reducción del riesgo y de fomento de la confianza.

Establecer a tiempo una gobernanza internacional adecuada en materia de IA, con la flexibilidad necesaria para responder a los rápidos avances tecnológicos, puede contribuir a evitar carreras armamentistas impulsadas por la tecnología y, al mismo tiempo, poner la innovación al servicio de la seguridad mundial.

Nueva Zelanda

[Original: inglés]
[11 de abril de 2025]

Esta comunicación nacional de Nueva Zelanda se presenta en respuesta a la nota verbal de fecha 12 de febrero de 2025 de la Oficina de Asuntos de Desarme y debe leerse junto con la respuesta de Nueva Zelanda a la nota verbal de la misma Oficina de fecha 1 de febrero de 2024².

Posición de Nueva Zelanda respecto de la inteligencia artificial en el ámbito militar

Nueva Zelanda reconoce que las aplicaciones potenciales y existentes de la inteligencia artificial (IA) en el ámbito militar tendrán repercusiones de gran alcance y múltiples dimensiones.

Por el momento, aunque no se sabe a ciencia cierta qué naturaleza y alcance tendrán muchas de esas repercusiones, hay organizaciones militares que ya están utilizando la IA en gran diversidad de funciones militares, como la inteligencia, la planificación, la logística, la navegación y la comunicación. Aunque entraña ciertos riesgos, la IA en el ámbito militar puede ofrecer importantes ventajas a los usuarios, como una mayor rapidez, eficacia, precisión y conciencia situacional. Al igual que otros ejércitos, las Fuerzas de Defensa de Nueva Zelanda pretenden aprovechar las oportunidades que ofrece la IA para mejorar sus operaciones y mantener la interoperabilidad con sus asociados.

Reiteramos el párrafo 1 de la resolución 79/239 de la Asamblea General, que reza como sigue: “que el derecho internacional, incluidos la Carta de las Naciones Unidas, el derecho internacional humanitario y el derecho internacional de los derechos humanos, se aplica a las cuestiones que por él se rigen y que surgen a lo largo de todas las etapas del ciclo de vida de la inteligencia artificial, incluidos los sistemas habilitados mediante la inteligencia artificial, en el ámbito militar”. Además de las obligaciones jurídicas vinculantes, deben tenerse en cuenta las normas éticas pertinentes durante todo el ciclo de vida de la IA en el ámbito militar.

Nueva Zelanda reconoce que la IA es de interés para el desarrollo y el uso de algunos sistemas de armas, por ejemplo para incrementar los niveles de autonomía. La posición de Nueva Zelanda sobre los sistemas de armas autónomos se detalla en su respuesta a la nota verbal de la Oficina de Asuntos de Desarme de fecha 1 de febrero de 2024.

Cabe imaginar que la IA pueda aplicarse al desarrollo de armas de destrucción masiva. Las armas biológicas y químicas están claramente prohibidas por el derecho internacional y Nueva Zelanda sostiene que, cuando se haga uso de la IA para desarrollar armas de ese tipo, se aplicará el criterio de finalidad general, tanto de la Convención sobre la Prohibición del Desarrollo, la Producción y el Almacenamiento de Armas Bacteriológicas (Biológicas) y Toxínicas y sobre Su Destrucción como de la Convención sobre la Prohibición del Desarrollo, la Producción, el Almacenamiento y el Empleo de Armas Químicas y sobre Su Destrucción, lo que significa, entre otras cosas, que la IA no debe utilizarse con ese fin. Además, como advirtieron varios Estados partes en el Tratado sobre la Prohibición de las Armas Nucleares, entre ellos Nueva Zelanda, es esencial que se mantenga un control humano determinante sobre

² Puede consultarse en www.mfat.govt.nz/assets/Peace-Rights-and-Security/Disarmament/New-Zealand-submission-to-the-UN-Secretary-General-on-autonomous-weapon-systems.pdf.

las armas nucleares y sus sistemas vectores mientras no se eliminen y se consiga un mundo libre de armas nucleares.

Propuestas normativas existentes y emergentes

Llegar a entendimientos comunes y establecer normas son tareas importantes para promover un uso militar responsable de la IA. En 2024, Nueva Zelanda se sumó, junto con muchos otros países, a la Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar promovida por los Estados Unidos, en la que se afirma que “el uso de IA en el ámbito militar puede y debería ser ético, responsable e incrementar la seguridad internacional”. Nueva Zelanda también ha participado en las Cumbres sobre la Inteligencia Artificial Responsable en el Ámbito Militar.

Nueva Zelanda valora positivamente los debates multilaterales sobre la elaboración y concertación de normas que regulen la IA en el ámbito militar organizados por las Naciones Unidas y otras entidades. Es importante que los interesados no estatales, como la sociedad civil, las organizaciones internacionales y regionales y el sector privado, participen en esos procesos.

Países Bajos (Reino de los)

[Original: inglés]
[7 de abril de 2025]

El Reino de los Países Bajos acoge con satisfacción la oportunidad de presentar sus opiniones sobre los retos y oportunidades para la paz y la seguridad internacionales que plantea la inteligencia artificial (IA) en el ámbito militar, en respuesta a la resolución [79/239](#), que la Asamblea General aprobó el 24 de diciembre de 2024.

Los Países Bajos reconocen las posibles aplicaciones militares de la IA y se comprometen a desarrollarla, desplegarla y utilizarla de forma responsable. La posición fundamental de los Países Bajos es que la aplicación de la IA en el ámbito militar debe ajustarse al derecho internacional, incluida la Carta de las Naciones Unidas, el derecho internacional humanitario y el derecho internacional de los derechos humanos.

Los días 15 y 16 de febrero de 2023, los Países Bajos acogieron la Primera Cumbre sobre la Inteligencia Artificial Responsable en el Ámbito Militar. Desde entonces, el proceso de Inteligencia Artificial Responsable en el Ámbito Militar ha ofrecido a representantes gubernamentales, centros de conocimiento, grupos de reflexión, empresas y organizaciones de la sociedad civil un foro multipartito en el que examinar las principales oportunidades y retos asociados a las aplicaciones militares de la IA. Cada año se celebra un debate de escala mundial y a lo largo del año y varios eventos regionales sobre Inteligencia Artificial Responsable en el Ámbito Militar, organizados hasta la fecha por Singapur, Kenya, Türkiye, Chile y los Países Bajos.

En la edición de la Cumbre que se celebró en 2023, los Países Bajos y otros 57 Estados suscribieron un llamamiento conjunto a la acción en materia de desarrollo, despliegue y uso responsables de la IA en el ámbito militar. En 2024, los Países Bajos respaldaron el Plan de Acción concertado en la Cumbre sobre la Inteligencia Artificial Responsable en el Ámbito Militar de ese año, que organizó la República de Corea y copatrocinaron los Países Bajos. El país también respaldó la Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar.

En la Cumbre de 2023, los Países Bajos pusieron en marcha la Comisión Global sobre Inteligencia Artificial Responsable en el Ámbito Militar y le encomendaron la tarea de formular recomendaciones aplicables a corto y largo plazo a los Gobiernos y, en general, a los distintos interesados. Los Países Bajos esperan que el informe de orientación estratégica de la Comisión se publique en septiembre de 2025.

En la sección siguiente se resume la posición de los Países Bajos y se exponen las cuestiones clave que requieren un examen más detenido.

Oportunidades para la paz y la seguridad internacionales

Desde una perspectiva militar, las principales ventajas de la IA son la rapidez y la escala. Esta tecnología permite procesar y analizar datos mucho más rápido. Los sistemas de elaboración de hipótesis y de apoyo para la toma de decisiones que funcionan con IA también ayudan a los mandos a plantear líneas de actuación, lo que mejora la visión estratégica y la capacidad de responder a las amenazas con prontitud y eficacia.

Los Países Bajos creen que la IA también puede contribuir a la paz y la seguridad internacionales al mejorar la comprensión, la conectividad y la protección de los civiles y reducir los riesgos en las operaciones de primera línea:

- Los sistemas de análisis y apoyo para la toma de decisiones que funcionan con IA mejoran la posición informativa de los mandos, en lo que respecta tanto a la situación sobre el terreno como a la evolución estratégica a largo plazo, lo que permite conocer mejor la dinámica de la población civil en zonas de conflicto, los problemas de la seguridad climática, los casos de violencia de género y las pautas de comportamiento de las organizaciones terroristas. A su vez, esa información puede servir para mejorar la gestión de riesgos y conflictos, contribuyendo así a la paz y la seguridad internacionales.
- Los Países Bajos consideran positiva la aplicación militar de IA para mejorar la conectividad entre fuerzas de defensa y entre estas y otros agentes, como las entidades de asistencia humanitaria, las organizaciones de vigilancia y las administraciones locales. Los datos pueden circular entre un gran número de usuarios, creando “fuentes únicas de datos” con sensores “inteligentes” que operen en un entorno protegido en red. También se puede recurrir a agentes de IA para transmitir datos a velocidades cada vez mayores. La mejora de la conectividad gracias a una comunicación de datos optimizada y más rápida redundará en beneficio de la paz y la seguridad internacionales, ya que mejora la comunicación, el intercambio de información y la cooperación internacional, por ejemplo en los sistemas de alerta temprana y la gestión de crisis.
- Los Países Bajos consideran que la IA tiene un gran potencial para la protección de la población civil. Esta tecnología permite reconocer patrones y desviaciones en grandes volúmenes de datos, posibilitando una comprensión más cabal del entorno civil, lo que, a su vez, puede reducir los riesgos de identificaciones erróneas, daños colaterales y bajas civiles. En general, la IA ofrece la posibilidad de mejorar la detección de posibles amenazas a civiles y bienes de carácter civil, haciendo que las fuerzas armadas puedan dar una respuesta rápida y adecuada. También puede contribuir a optimizar labores de asistencia humanitaria, como el suministro de alimentos, refugio y atención médica en zonas de conflicto. Por último, la IA puede mejorar las investigaciones sobre las bajas civiles mediante la recopilación y el análisis de datos y pruebas con los que determinar la causa del daño y garantizar que los responsables respondan por sus actos.
- La IA reduce los riesgos para el personal militar de primera línea, ya que los sistemas autónomos que utilizan esta tecnología pueden sustituir a los humanos

en determinadas actividades en terrenos difíciles o peligrosos. Ejemplo de ello son la vigilancia submarina y el apoyo a las operaciones de búsqueda y salvamento en condiciones meteorológicas extremas. La IA también puede ayudar a reducir los gastos médicos y de rehabilitación al limitar la exposición del personal militar a entornos de alto riesgo.

Retos para la paz y la seguridad internacionales

Los Países Bajos constatan varios riesgos para la paz y la seguridad internacionales derivados de la aplicación de la IA en el ámbito militar:

- A los Países Bajos les preocupa que la IA pueda utilizarse para amplificar, mejorar y automatizar los ciberataques y la manipulación de la información, que socavan la paz y la seguridad internacionales. El auge de la IA generativa ha hecho que manipular información y organizar ciberataques automatizados resulte más fácil. Cuando se llevan a cabo en el ámbito militar, este tipo de acciones perturban las líneas de comunicación operacional y complican la toma de decisiones. A largo plazo, la difusión generalizada de desinformación y los ciberataques automatizados pueden erosionar la confianza en las líneas de comunicación militar. También pueden minar la confianza entre Estados y deteriorar consiguientemente relaciones frágiles, sobre todo entre países que ya están al borde de un posible conflicto.
- Los riesgos asociados a la aplicación de la IA en el ámbito militar pueden dar lugar a sistemas que vulneren el derecho internacional. Estas deficiencias podrían deberse a una adaptación insuficiente al contexto, a los datos y a la jerga militar y, a su vez, llevar a simplificar en exceso la toma de decisiones militares o a desatender contextos operacionales específicos, por ejemplo. Los Estados también podrían infringir obligaciones jurídicas internacionales si una aplicación se comporta de forma imprevista, genera resultados discriminatorios basados en características irrelevantes o propone líneas de actuación ilícitas. Debido a la creciente prevalencia de la IA, el sesgo de automatización, los sesgos en los conjuntos de datos y las decisiones humanas basadas en sistemas inadecuados de IA podrían dificultar considerablemente la atribución de responsabilidades, así como la rendición de cuentas y las medidas de reparación adecuadas. Hay que tener en cuenta que no cabe esperar que las aplicaciones de IA razonen o funcionen del mismo modo que los humanos.
- El riesgo de intensificación de tensiones inducido por la IA puede poner en peligro la paz y la seguridad internacionales. La capacidad de operar más rápido y con magnitudes más grandes de la IA está haciendo que se acelere el bucle “observar-orientar-decidir-actuar”, por lo que las discrepancias entre las intenciones militares y los análisis generados por sistemas que funcionan con IA pueden dar lugar a percepciones erróneas, con lo que existe la posibilidad de que la IA contribuya accidentalmente a intensificar tensiones. Los sistemas de IA son capaces de identificar más rápido mayor número de objetivos potenciales, por lo que utilizarla puede hacer que también aumente la intensidad y la letalidad de los conflictos.
- En consecuencia, crear sistemas defensivos sólidos resulta cada vez más difícil. La velocidad a la que surgen las nuevas aplicaciones de IA dificulta la adopción de estrategias y tácticas para contrarrestarlas y defenderse de ellas eficazmente en un contexto militar. Esta consecuencia específica del uso creciente de los sistemas de IA puede propiciar acciones ofensivas y, por ende, tener repercusiones negativas para la paz y la seguridad internacionales.

- Con el acceso de organizaciones terroristas, redes de delincuencia organizada y otros actores no estatales a capacidades militares de IA, la desestabilización se ha convertido en un nuevo motivo de preocupación. En estas circunstancias, a los Países Bajos les preocupa que la IA pueda facilitar a estos actores la posibilidad de fabricar armas químicas, biológicas, radiológicas y nucleares.

Dada la rápida evolución de las tecnologías de IA, los Países Bajos son conscientes de que actualmente no es posible prever todos los retos y oportunidades que acarrearán para la paz y la seguridad internacionales. Algunos son totalmente nuevos y otros ya existen, pero pueden verse agudizados por la aplicación de esta tecnología. Es esencial mantener un diálogo internacional permanente sobre esta cuestión para lograr que todos los Estados apliquen de forma responsable la IA en el ámbito militar.

Aplicación responsable de la inteligencia artificial en el ámbito militar

Para que la IA se aplique de forma responsable en el ámbito militar, deben mantenerse un juicio y un control humanos que tengan en cuenta el contexto. Los humanos deben seguir siendo responsables y rendir cuentas. Con todo, es importante tomar nota de los puntos que se exponen a continuación.

Un mayor control humano no garantiza una inteligencia artificial más responsable

Los Países Bajos consideran que no existe un único método de integrar un nivel suficiente de juicio y control humanos en las aplicaciones de IA. El grado en que se ejercen estas capacidades es muy variable: dependiendo de diversos factores, pueden ejercerse directamente o coexistir con niveles elevados de automatización y autonomía, por lo que ha de decidirse caso por caso. Ésta es la única forma de tener en cuenta múltiples factores, como el contexto operacional, el impacto en la capacidad de la tecnología de operar de forma autónoma en entornos hostiles, los parámetros del sistema y la interacción hombre-máquina.

La investigación y desarrollo es clave para el despliegue responsable de aplicaciones de inteligencia artificial en el ámbito militar

Los Países Bajos consideran que la investigación y el desarrollo son importantes. Los Estados deben evaluar adecuadamente si sus aplicaciones de IA funcionan de la manera para la que se concibieron y si pueden desplegarse en un contexto de uso específico. Esto es especialmente necesario en situaciones de combate y otros entornos de alto riesgo. La investigación y desarrollo en general y los procedimientos probados y fiables de ensayo, evaluación, verificación y validación de aplicaciones concretas de IA permiten detectar posibles problemas y subsanarlos o mitigarlos antes de desplegarlas. Además, es importante que el personal militar reciba la capacitación adecuada sobre las aplicaciones de IA y se familiarice con ellas antes de que se desplieguen, para asegurarse de que comprenden sus capacidades y limitaciones, algo especialmente importante si se tiene en cuenta que la tecnología de esas aplicaciones progresó cada vez más rápido y que utilizarlas es cada vez menos costoso.

La gobernanza internacional de la inteligencia artificial militar debe ser flexible, inclusiva y realista

Por lo que respecta a la gobernanza internacional de la IA en el ámbito militar, los Países Bajos reconocen la necesidad de un enfoque flexible, equilibrado y realista. En primer lugar, los marcos de gobernanza deben ser flexibles para seguir la rápida evolución tecnológica y del campo de batalla. En segundo lugar, las partes deben trabajar para llegar a una comprensión común de la IA en el ámbito militar y de las

oportunidades, riesgos y posibles soluciones que conlleva. Esto requerirá un diálogo global inclusivo y la participación activa de todos los grupos interesados, incluidos los Estados, los centros de conocimiento, la sociedad civil y la industria. En tercer lugar, los Estados deben centrarse en establecer salvaguardias para una aplicación responsable de la IA en el ámbito militar, por ejemplo, ocupándose de cuestiones como la trazabilidad o la comprensibilidad. En cuarto lugar, la gobernanza internacional del despliegue militar de la IA debe tener en cuenta los diferentes puntos de vista en materia de regulación de los Estados. En los límites de las obligaciones legales vigentes, la gobernanza internacional de la IA en el ámbito militar no debe entorpecer la capacidad de innovación de los Estados.

Debate sobre los sistemas de armas autónomos

Habida cuenta del gran potencial de la IA para el manejo de sistemas de armas autónomos, existen claros paralelismos entre el debate general sobre el uso de esta tecnología en el ámbito militar y el debate sobre la regulación de los sistemas de armas autónomos. Los Países Bajos consideran que los debates internacionales sobre ambos temas son complementarios y mutuamente beneficiosos.

Pakistán

[Original: inglés]
[9 de abril de 2025]

El rápido avance y la integración de las tecnologías de inteligencia artificial (IA) en el ámbito militar están a punto de transformar radicalmente la guerra. La IA está cada vez más integrada en las operaciones militares mediante su aplicación en sistemas de armas autónomos, en las funciones de mando y control, en sistemas de apoyo para la toma de decisiones y en los ámbitos de la inteligencia, la vigilancia y el reconocimiento, el adiestramiento, la logística y la guerra cibernetica y de la información. Aunque estos avances favorecen la eficiencia operacional, también plantean riesgos considerables para la paz y la seguridad internacionales.

Retos asociados a la inteligencia artificial en el ámbito militar

Riesgos estratégicos: interacción con las armas nucleares

La integración de la IA en los sistemas de armas nucleares plantea riesgos estratégicos, especialmente en el mando, el control y las comunicaciones nucleares. Cuando las capacidades de la IA se combinan con las políticas relativas al dispositivo militar nuclear, así como a su empleo, pueden dar lugar a errores de cálculo, accidentes y consecuencias catastróficas.

El concepto de disuisión nuclear depende en gran medida de la racionalidad, la percepción y la toma de decisiones políticas de los seres humanos. La integración de la IA puede eliminar o reducir considerablemente estos factores humanos críticos, lo que aumenta el riesgo de escaladas automatizadas o accidentales. Reconociendo estas profundas preocupaciones, algunos Estados se han comprometido públicamente a mantener un control humano determinante sobre las decisiones de empleo de armas nucleares, un principio que el Pakistán apoya e insta a todos los Estados poseedores de armas nucleares a respaldar.

En las regiones con armas nucleares, la dependencia de los sistemas de apoyo para la toma de decisiones basados en la IA y de los sistemas de armas totalmente autónomos en la esfera convencional también puede generar riesgos de escalada. Eliminar por completo el control humano durante las crisis podría dificultar el control de la magnitud y la duración de los conflictos. La automatización de las respuestas en

situaciones volátiles y de alto riesgo, especialmente en regiones con tensas dinámicas nucleares, puede agravar el entramado convencional-nuclear y afectar negativamente a la estabilidad estratégica.

El uso de la IA para la evaluación de datos y en actividades de inteligencia, vigilancia y reconocimiento puede generar una falsa sensación de confianza para los Estados que consideren la posibilidad de realizar ataques preventivos y desestabilizadores de contrafuerza o establecer como objetivo las capacidades de contraataque, lo que plantea graves riesgos para la estabilidad regional y mundial.

Riesgos operacionales – Pérdida de capacidad de acción humana

La autonomía de las operaciones militares impulsada por la IA puede reducir la supervisión humana y complicar la gestión de las crisis. A medida que la guerra va asumiendo la “velocidad de las máquinas”, la toma de decisiones humanas se ve gravemente constreñida, lo que reduce las posibilidades de mitigar las crisis y actuar por vía diplomática.

Los seres humanos pueden confiar excesivamente en recomendaciones de los sistemas de apoyo para la toma de decisiones que hayan sido generadas por IA, incluso si son erróneas o incompletas, lo que da lugar a un sesgo de automatización. Las decisiones militares críticas podrían depender excesivamente de los resultados de las máquinas, lo que llevaría a los mandos a pasar por alto la intuición humana, el contexto o la prudencia, con el riesgo de agravar involuntariamente los conflictos.

Las capacidades potenciadas por la IA, impulsadas por el atractivo de una mayor eficiencia operacional y la carrera por obtener una ventaja decisiva, podrían provocar una mayor predisposición a su uso, lo que reduciría el umbral para los conflictos armados. En tiempos de crisis, un umbral bajo para el uso de la fuerza sería altamente desestabilizador.

Riesgos técnicos

Las aplicaciones militares de la IA pueden entrañar vulnerabilidades técnicas, como sesgos algorítmicos, envenenamiento de datos y susceptibilidad a los ciberataques. Podrían surgir conflictos debido al mal funcionamiento o la manipulación de los sistemas de alerta temprana o a ataques de envenenamiento de datos. Las capacidades de la IA suelen funcionar como “cajas negras”, ya que generan decisiones que carecen de transparencia o explicabilidad, lo que complica la validación y la rendición de cuentas. Estas vulnerabilidades pueden provocar resultados impredecibles, fallos del sistema y riesgos significativos para la integridad operacional. Las capacidades de la IA probadas en un entorno con conjuntos de datos específicos pueden no funcionar de manera fiable en entornos completamente diferentes con dinámicas más complejas.

Riesgos normativos, jurídicos y éticos

El uso de la IA en el ámbito militar plantea retos éticos, normativos y jurídicos, sobre todo en lo que respecta a la observancia del derecho internacional humanitario, que se basa fundamentalmente en el juicio humano, el discernimiento y la toma de decisiones en función del contexto, cualidades que, por su propia naturaleza, son difícilmente replicables por los sistemas de IA. Delegar en sistemas autónomos funciones críticas como la selección de objetivos y el combate, incluidas las decisiones sobre el uso de la fuerza letal, entraña el riesgo de violar principios fundamentales del derecho internacional humanitario, como los de distinción, proporcionalidad, precauciones en el ataque y necesidad militar. Los sistemas de IA que producen resultados impredecibles, poco fiables o inexplicables complican aún

más la observancia del derecho internacional humanitario, pudiendo provocar daños ilícitos o involuntarios.

Además, el hecho de que el ser humano no intervenga directamente en la toma de decisiones o de que se dependa en exceso de sistemas de apoyo para la toma de decisiones basados en la IA plantea importantes interrogantes en materia de rendición de cuentas y responsabilidad, lo que dificulta enormemente la imputación y la exigencia de responsabilidades cuando se cometan actos ilícitos o indebidos. Si algo sale mal, los mandos podrían descargar la responsabilidad en la IA, lo que complicaría la rendición de cuentas ante la justicia y las investigaciones de posibles crímenes de guerra.

Delegar decisiones de vida o muerte en sistemas autónomos plantea además cuestiones éticas, ya que puede ir en detrimento de la compasión, el razonamiento moral y el juicio humano y agravar así el riesgo de violencia injustificada y bajas civiles.

Proliferación y riesgos para la seguridad mundial

La proliferación de tecnologías militares de IA plantea importantes riesgos para la seguridad internacional. La difusión de capacidades avanzadas de IA, en particular las armas autónomas, podría impulsar nuevas carreras armamentistas y desestabilizar los entornos de seguridad regionales y mundiales. La facilidad con la que proliferan y pueden ser adquiridas por agentes no estatales hacen que la cuestión sea aún más preocupante.

Propuesta de respuesta internacional: el papel central de los mecanismos de las Naciones Unidas

Las tecnologías de IA son de uso general y sus usos pacíficos son esenciales para la consecución de los Objetivos de Desarrollo Sostenible. Al mismo tiempo, las implicaciones de la IA en el ámbito militar son transversales y pueden afectar significativamente a la paz y la seguridad internacionales, por lo que requieren una respuesta internacional coordinada.

El Pakistán reconoce el valor de las iniciativas de gobernanza de la IA que se emprenden fuera del ámbito de las Naciones Unidas, pero es consciente de sus limitaciones, en particular en lo que respecta a la participación universal y la legitimidad multilateral formal. Si bien estas iniciativas pueden complementar los esfuerzos de las Naciones Unidas al fomentar el diálogo y la voluntad política, llevarlas a cabo de manera aislada conlleva el riesgo de fragmentación. Por lo tanto, los debates sobre las aplicaciones militares de la IA deberían trasladarse a los foros de las Naciones Unidas en aras de la inclusividad y la legitimidad y de un marco global coherente que refleje los intereses de todos los Estados.

Por estas razones, las Naciones Unidas deben seguir estando en el centro de cualquier respuesta internacional. El mecanismo de desarme de las Naciones Unidas debería desempeñar un papel central en el desarrollo de un marco de gobernanza internacional para la IA militar y en la prevención de la fragmentación del panorama normativo. La magnitud y la novedad de las implicaciones militares de la IA requieren una respuesta multidimensional, holística y multilateral. Gracias a su composición universal, las Naciones Unidas constituyen un foro ideal en el que todos los Estados, desarrollados y en desarrollo, tienen voz.

Pero no bastará con un único foro o instrumento. Se necesita una estrategia estructurada que utilice múltiples órganos de desarme de las Naciones Unidas, en la que cada foro aborde la cuestión desde su perspectiva y mandato específicos, de manera complementaria. Proponemos aprovechar todos los foros pertinentes, desde la Asamblea General y su Primera Comisión hasta la Comisión de Desarme, la Conferencia de Desarme y la Convención sobre Prohibiciones o Restricciones del

Empleo de Ciertas Armas Convencionales que Puedan Considerarse Excesivamente Nocivas o de Efectos Indiscriminados. Así podrían abordarse de forma global las dimensiones estratégica, humanitaria, jurídica y técnica, evitando lagunas y redundancias. El trabajo de cada foro debería enriquecer el los demás, creando sinergias hacia el objetivo común de mitigar los riesgos de la IA militar y, al mismo tiempo, preservar el uso pacífico de la IA.

Conferencia de Desarme

La Conferencia de Desarme debería dar prioridad a la cuestión de los riesgos estratégicos asociados a la IA militar, particularmente en el ámbito nuclear, en consonancia directa con los temas 1 y 2 de su agenda (Cesación de la carrera de armamentos nucleares y desarme nuclear y Prevención de la guerra nuclear, incluidas todas las cuestiones conexas). En 2023, el Pakistán propuso establecer un nuevo tema de la agenda de la Conferencia sobre esta cuestión ([CD/2334](#)).

En el marco de ese nuevo tema de la agenda, la Conferencia de Desarme debería establecer un órgano subsidiario o un grupo *ad hoc* con el mandato específico de examinar los riesgos que entraña la IA militar para la estabilidad, evaluar su contribución a los riesgos nucleares y proseguir las negociaciones sobre medidas concretas, entre las que cabría incluir las siguientes:

- El compromiso de mantener el control humano y no sustituir el juicio humano en las decisiones relativas al empleo de armas nucleares;
- La prohibición de utilizar capacidades de IA para manipular datos o atacar sistemas de mando, control y comunicaciones nucleares; y
- El desarrollo de medidas para limitar el despliegue y el uso de determinadas capacidades de la IA que pueden desencadenar ataques anticipatorios y contribuir al recrudecimiento de los riesgos nucleares.

La Conferencia de Desarme está especialmente indicada para esos debates, ya que congrega a todos los Estados de importancia militar en pie de igualdad y funciona por consenso, salvaguardando así los intereses vitales de todos los Estados en materia de seguridad. Abordar esta cuestión permitiría revitalizar la labor de la Conferencia de Desarme, que mostraría así su capacidad de respuesta ante amenazas nuevas e incipientes.

Comisión de Desarme

Con su composición universal y su mandato deliberativo, la Comisión de Desarme se encuentra en una posición ideal para elaborar directrices y recomendaciones prácticas sobre el uso responsable de la IA con fines militares. A lo largo de su historia, la Comisión de Desarme ha elaborado con eficacia directrices similares (por ejemplo, medidas de fomento de la confianza en 1988 y enfoques regionales del desarme en 1993).

En su Grupo de Trabajo II, la Comisión de Desarme podría elaborar directrices y recomendaciones sobre medidas destinadas a fomentar la confianza y la seguridad en relación con las aplicaciones militares de la IA, tanto a nivel mundial como regional. Entre sus elementos clave podrían figurar la reafirmación de los fundamentos normativos, la recomendación de medidas operacionales y técnicas para mitigar los riesgos, la elaboración de estrategias de reducción de los riesgos de la IA militar y la solución de problemas de proliferación, garantizando al mismo tiempo el acceso equitativo a los usos pacíficos de la IA.

Primera Comisión de la Asamblea General de las Naciones Unidas

La Primera Comisión de la Asamblea General debería instituir la presentación periódica de informes de evaluación del Secretario General de las Naciones Unidas y llevar un registro del desarrollo tecnológico de capacidades militares de IA y de los riesgos asociados a partir de la información que faciliten voluntariamente los Estados Miembros. Esas evaluaciones periódicas ofrecerían una visión fidedigna de la evolución de las capacidades, aportarían información oportuna y facilitarían la formulación de respuestas de política internacional bien fundamentadas.

Al examinar esos informes, la Primera Comisión podría organizar debates específicos sobre IA y tal vez establecer un grupo de trabajo de composición abierta en el marco de la Asamblea General para negociar una plataforma más institucional, por ejemplo, un registro de las Naciones Unidas sobre las aplicaciones militares de la IA (aunque, por el momento, es preferible aprovechar los foros existentes).

Los informes también podrían servir para señalar las esferas en que se está llegando a un consenso o en que es necesario seguir trabajando y para orientar los programas de foros como la Conferencia de Desarme, la Comisión de Desarme y la Convención sobre Ciertas Armas Convencionales.

Convención sobre Ciertas Armas Convencionales

El Grupo de Expertos Gubernamentales de la Convención sobre Ciertas Armas Convencionales sigue siendo esencial para abordar las implicaciones humanitarias, éticas y jurídicas de los sistemas de armas autónomos letales. Su carácter inclusivo (con la participación de la sociedad civil y del Comité Internacional de la Cruz Roja como observadores) es una ventaja.

El Pakistán valora el trabajo que dicho Grupo viene realizando desde 2017, especialmente los 11 principios rectores establecidos en 2019. Sin embargo, los avances de la Convención han sido lentos y se han basado sobre todo en principios, en lugar de centrarse en reglamentaciones concretas. El Pakistán comparte el análisis de que, en sus deliberaciones, la Convención ha prestado una “atención insuficiente y decreciente” a las dimensiones de seguridad de las armas potenciadas por la IA, lo que pone de relieve la necesidad de que ese y otros foros tomen medidas adicionales. No obstante, en el ámbito humanitario, el Grupo de Expertos Gubernamentales de la Convención debería proseguir e intensificar su labor.

El Pakistán aboga por que se concluyan las negociaciones sobre un protocolo jurídicamente vinculante de la Convención que prohíba los sistemas de armas autónomos letales que funcionen sin control humano o no puedan atenerse al derecho internacional humanitario. El mandato actual del Grupo de Expertos Gubernamentales permite a los Estados Miembros elaborar elementos de dicho instrumento para presentarlos en la Séptima Conferencia de Examen de las Altas Partes Contratantes en la Convención y poder iniciar negociaciones oficiales posteriormente.

Conclusión

El Pakistán pone de relieve la necesidad de una acción internacional coordinada e inclusiva para mitigar importantes riesgos de la IA militar. Propone un enfoque de gobernanza que equilibra la seguridad y el desarrollo, en la medida en que garantiza la estabilidad y, al mismo tiempo, permite un desarrollo beneficioso de la IA. Mediante una estrategia estructurada que abarque múltiples foros dentro de las Naciones Unidas, la comunidad internacional puede establecer sólidas salvaguardas normativas, mantener la seguridad internacional y preservar el acceso equitativo y no discriminatorio a los usos pacíficos de la IA.

Perú

[Original: español]
 [11 de abril de 2025]

La resolución [79/239](#) de la Asamblea General de las Naciones Unidas, aprobada el 24 de diciembre de 2024 con el voto favorable del Perú, solicita al Secretario General, en su párrafo 7, que recabe:

las opiniones de los Estados Miembros y de los Estados observadores sobre las oportunidades y los retos que plantea para la paz y la seguridad internacionales la aplicación de la inteligencia artificial en el ámbito militar, con atención específica a los ámbitos distintos de los sistemas de armas autónomos letales, y que, en su octogésimo período de sesiones, le presente un informe sustantivo en el que se resuman esas opiniones y se clasifiquen las propuestas normativas existentes y emergentes, con un anexo en el que figuren esas opiniones, para que los Estados sigan debatiendo.

En atención a ello, el Perú presenta a continuación algunos elementos de su posición con el fin de contribuir a la elaboración del mencionado informe del Secretario General.

I. Relevancia de la inteligencia artificial en el ámbito militar

El Perú reconoce la rápida y dinámica evolución de las tecnologías emergentes en el ámbito militar, en particular, las potenciales aplicaciones de la inteligencia artificial (IA). Si bien se observa con atención los desarrollos en este campo, incluyendo formas en las que la IA parece estar transformando las operaciones militares, desde el uso de drones autónomos hasta los sistemas de apoyo a la toma de decisiones, se considera indispensable promover un diálogo multilateral sostenido orientado a establecer principios que garanticen un uso ético y responsable de dichas herramientas.

Dado que la IA puede integrarse tanto en sistemas de armas como en sistemas de apoyo a las operaciones militares, el Perú estima fundamental abordar los desafíos y preocupaciones que plantea su uso desde las perspectivas humanitaria, jurídica, de seguridad, tecnológica y ética, incluyendo los riesgos vinculados a sesgos algorítmicos. Estas inquietudes se ven acentuadas por las implicancias que el empleo de esta tecnología podría tener para la estabilidad y la seguridad internacional.

Ello es aún más preocupante frente al riesgo de las implicaciones del uso de la IA en relación con armas nucleares y otras armas de destrucción masiva. Ante ello, es imperativo reforzar el principio de control humano significativo.

II. Opiniones

Cumplimiento del derecho internacional

El desarrollo, la implementación y la utilización de tecnologías basadas en IA en el ámbito militar deben cumplir plenamente con el derecho internacional, incluyendo el derecho internacional de los derechos humanos y el derecho internacional humanitario, así como con los principios fundamentales recogidos en la Carta de las Naciones Unidas.

En tal sentido, todo desarrollo normativo para regular la IA en el ámbito militar deberá garantizar su uso responsable y ético, que garantice también la no proliferación de tecnologías militares basadas en la IA y el acceso equitativo al conocimiento y capacidades tecnológicas.

Con ello, se busca asegurar que cualquier uso de la IA respete la dignidad humana, proteja a la población civil y garantice la estabilidad y la paz internacionales.

Reconocimiento de los beneficios y riesgos

La IA ofrece valiosas oportunidades para una mayor comprensión de las situaciones operativas y, por ende, para mejorar la aplicación del derecho internacional humanitario y la protección de los civiles y bienes de carácter civil.

No obstante, su empleo puede acarrear riesgos previsibles e imprevisibles en el ámbito militar, por ejemplo, a raíz de sesgos algorítmicos, fallos de diseño, uso indebido o malicioso, entre otros. Asimismo, la IA puede repercutir en dinámicas regionales y globales complejas, al impactar los riesgos de escalada, error de cálculo, reducción del umbral de conflicto o el surgimiento de una carrera armamentista.

Desarrollo responsable

El uso de la IA en el ámbito militar debe promover la paz y la protección de la población civil, y que los avances tecnológicos deben complementar, no sustituir, las capacidades humanas.

En concordancia con los principios aplicables a los sistemas de armas autónomas, la aplicación de la IA en el ámbito militar debe asegurar que la responsabilidad y la rendición de cuentas nunca puedan ser transferidas a las máquinas. En ese sentido, se subraya la necesidad de preservar un control humano significativo en toda decisión que implique el uso de la fuerza.

Se mantiene que todos los riesgos y desafíos relacionados con esta tecnología deben abordarse de manera integral durante todo su ciclo de vida.

El establecimiento de controles y salvaguardias para prevenir un uso indebido de esta tecnología en el ámbito militar puede llevarse a cabo sin obstaculizar la investigación, el desarrollo, la experimentación ni la innovación vinculadas con la IA en otros campos.

Implementación y transparencia

Resulta prioritaria la definición de estrategias, principios, estándares y normas, así como políticas y marcos legales nacionales que garanticen un uso responsable de la IA en el ámbito militar.

Igualmente, se resalta la importancia de establecer medidas de fomento de la confianza y de reducción de riesgos, así como mecanismos de intercambio de buenas prácticas, en aras de la transparencia y la cooperación entre los Estados.

Formato de discusiones

Se considera esencial mantener un diálogo continuo a nivel global, regional e interestatal sobre el desarrollo de medidas para garantizar una IA responsable en el ámbito militar.

Asimismo, se defiende una participación inclusiva en esta temática, que incluya las posiciones de los Estados, en particular de los Estados en desarrollo, así como los aportes de otras partes interesadas, como la industria, la academia, la sociedad civil y las organizaciones regionales e internacionales.

Se considera importante tomar en consideración que diferentes Estados y regiones se encuentran en distintas fases de integración de las capacidades de la IA en el ámbito militar y proceden de distintos entornos de seguridad.

En este contexto, se resalta el valor de impulsar la creación de capacidades en países en desarrollo y reforzar la cooperación internacional, con el objeto de reducir las brechas existentes y fortalecer su participación en los debates sobre el uso de esta tecnología.

Participación del Perú en discusiones internacionales

Cumbres sobre Inteligencia Artificial Responsable en el Dominio Militar

- El Perú participó en las ediciones de 2023 y 2024 y en su taller regional
- Adhesión a la declaración final de la Cumbre sobre Inteligencia Artificial Responsable en el Dominio Militar en el año 2024 (“Blueprint for Action”)

Declaración política sobre el uso militar responsable de la inteligencia artificial y la autonomía

- El Perú participó como observador en la reunión plenaria inaugural de esta iniciativa y posteriormente formalizó su adhesión

Cumbre de acción sobre inteligencia artificial – Military Talks (París, 2025)

- El Perú participó a alto nivel y se suscribió a la Declaración de París sobre el mantenimiento del control humano en sistemas de armas basados en la IA.

Reino Unido de Gran Bretaña e Irlanda del Norte

[Original: inglés]
[11 de abril de 2025]

La inteligencia artificial (IA) es una familia de tecnologías de uso general, cualquiera de las cuales puede permitir a las máquinas realizar tareas que tradicionalmente habrían requerido inteligencia humana o biológica, especialmente cuando las máquinas aprenden a partir de datos cómo realizar esas tareas. Las tecnologías de IA están madurando y adoptándose a un ritmo extraordinario. Al ser un conjunto de tecnologías con distintos sistemas, métodos y aplicaciones, sus trayectorias de desarrollo y sus consecuencias son también diferentes. Lo que es seguro es que tienen el potencial de impulsar un cambio capaz de transformar todos los aspectos de la sociedad, la economía y la política, incluidas la defensa y la seguridad.

El Reino Unido celebra que la resolución [79/239](#) de la Asamblea General le brinde la ocasión de examinar las implicaciones en el ámbito militar de aplicaciones de la IA distintas de las asociadas a los sistemas de armas autónomas letales, que han sido objeto de amplios y útiles debates, como los que está manteniendo el Grupo de Expertos Gubernamentales establecido en virtud de la Convención sobre Prohibiciones o Restricciones del Empleo de Ciertas Armas Convencionales que Puedan Considerarse Excesivamente Nocivas o de Efectos Indiscriminados. Una evaluación rigurosa de las implicaciones estratégicas globales de la IA militar, que reúna las reflexiones, ideas y buenas prácticas examinadas en foros internacionales oficiales y oficiosos sobre la cuestión, permitirá abordar de manera holística cómo aprovechar al máximo las oportunidades que encierra la IA en el ámbito militar, al tiempo que se abordan eficazmente los riesgos conexos.

Oportunidades de la inteligencia artificial en el ámbito militar

La integración de la IA en el ámbito militar podría transformar la defensa, la dinámica de la seguridad mundial y la naturaleza de la guerra. Las tecnologías avanzadas que funcionan con IA pueden categorizar y precisar grandes cantidades de

datos de distintas fuentes de forma más rápida y exhaustiva, por lo que permitirán ganar en eficacia, mejorar toma de decisiones y planificar operaciones en menos tiempo y de manera más rigurosa. La IA en los sistemas de inteligencia, vigilancia y reconocimiento puede ofrecer una imagen más precisa del contexto operacional y permitir a los responsables de la planificación que las operaciones afecten en menor medida a los civiles, lo que redundará en una mayor protección de la población y las infraestructuras civiles. Las funciones autónomas relativas a la logística y las municiones sin detonar harán que se necesite menos personal militar sobre el terreno. Por consiguiente, la IA en el ámbito militar podría reforzar la seguridad nacional e internacional y reducir el riesgo para la vida humana y el número de bajas.

En un estudio sobre la IA y el mantenimiento de la paz realizado por el Ministerio de Defensa del Reino Unido se señalan diversos beneficios que las capacidades y sistemas mejorados con IA podrían aportar a las operaciones de paz, como los siguientes:

- Una capacidad analítica que mejorará la conciencia situacional, la toma de decisiones operacionales, la planificación de posibles situaciones y la capacidad de análisis de los estados de opinión.
- Sistemas autónomos, como los vehículos aéreos no tripulados, con los que podrían cubrirse zonas geográficas más vastas o regiones de alto riesgo (donde puede resultar peligroso que el personal de mantenimiento de la paz tenga una presencia permanente).
- La logística podría mejorar la prestación de asistencia sanitaria y ayuda a las poblaciones locales, lo que contribuiría a los objetivos de la misión y fomentaría la confianza de la comunidad.

Estas capacidades pueden aplicarse para mejorar la supervisión y verificación de acuerdos de control de armamentos y acuerdos de paz, con lo que sería más fácil detectar infracciones o confirmar de manera oportuna y creíble el cumplimiento de esos acuerdos. Las herramientas de IA podrían permitir una mejor detección, identificación, atribución y verificación de diversos tipos de operaciones hostiles que no llegan a la categoría de conflictos, lo que reduciría la eficacia de tales actividades e incluso evitar su comisión. También pueden ayudar a vigilar y detectar en tiempo real discursos de odio en línea, campañas propagandísticas o cambios en la opinión pública que podrían agravar las tensiones o socavar la negociación de acuerdos de paz o de alto el fuego.

Desafíos y riesgos

El uso de la IA en el ámbito militar puede agudizar riesgos existentes y generar nuevas amenazas de intensidad superior o inferior a un conflicto armado. La adopción precipitada de capacidades de IA para obtener ventajas estratégicas podría llevar a los países a utilizar modalidades de IA inaceptables desde el punto de vista jurídico, ético o de seguridad. Los nuevos riesgos de intensificación de tensiones y los accidentes generados por la IA y causados por fallos de funcionamiento o por la fragilidad, debilidad, inmadurez o inseguridad de los sistemas de IA requerirán nuevos protocolos y mecanismos de distensión. Los actores hostiles pueden tratar de atacar los sistemas nacionales de IA y socavar la confianza en su desempeño, seguridad y fiabilidad (por ejemplo, “envenenando” fuentes de datos, corrompiendo componentes de equipos de cadenas de suministro o interfiriendo en comunicaciones y órdenes), lo que podría alterar los sistemas y sesgar la toma de decisiones militares en tiempos de crisis y otros entornos operacionales.

En tiempos de conflicto, puede que estas tecnologías —y el ritmo operacional que generan— condensen drásticamente los plazos de decisión, pongan a prueba los

límites de la comprensión humana y exijan responder a velocidad de máquina. La naturaleza de caja negra de muchas capacidades de IA hace que los humanos sean a menudo incapaces de discernir cómo o por qué se ha obtenido un resultado concreto. Las operaciones realizadas con IA pueden generar comportamientos impredecibles e incomprensibles y hacer que resulte difícil inferir o evaluar con precisión la intención de un adversario, dar lugar a interpretaciones erróneas o provocar consecuencias indeseadas. Los operadores podrían depositar una confianza excesiva en los resultados algorítmicos sin un conocimiento completo de los supuestos subyacentes, las limitaciones y los defectos de los sistemas de IA. A falta de salvaguardias, normas y protocolos adecuados, los sistemas que funcionan con IA podrían agudizar el riesgo de malentendidos, errores de cálculo y escaladas involuntarias.

La disponibilidad generalizada de capacidades o herramientas avanzadas de IA y otras tecnologías de doble uso probablemente aumente los riesgos de proliferación y desarrollo de armas novedosas por parte de actores estatales y no estatales. La IA también podría utilizarse para aumentar o avanzar en las actividades de desinformación dirigidas a generar hostilidad hacia ciertos países, lo que podría provocar conflictos e intensificar las tensiones.

Compromiso del Reino Unido con una inteligencia artificial segura y responsable en el ámbito militar

El Reino Unido es consciente de que la IA suscita hondas preocupaciones respecto de la imparcialidad, los sesgos, la fiabilidad y la naturaleza de la responsabilidad y la rendición de cuentas humanas, especialmente en contextos militares. Aunque los Estados tienen una larga tradición de incorporación de nuevas tecnologías y seguirán confiando en los sistemas jurídicos, de seguridad y regulación establecidos mucho tiempo atrás, debemos reconocer los problemas específicos que, por su naturaleza, plantea la IA y la importancia de mostrar de manera fehaciente que somos responsables y dignos de confianza.

El Reino Unido ha plasmado su compromiso con una IA segura y responsable en su estrategia de IA en el ámbito de la defensa y principios éticos asociados. Esos principios éticos sobre IA, recogidos en la política “ambiciosa, segura, responsable” del Reino Unido, establecen como consideraciones éticas generales la centralidad humana, la responsabilidad, la comprensión, la reducción de sesgos y daños y la fiabilidad. La publicación del servicio conjunto “Dependable artificial intelligence (AI) in defence”, presentada en noviembre de 2024, ofrece a los equipos del Ministerio de Defensa y otras entidades una orientación clara sobre cómo aplicar esos principios éticos para ofrecer servicios y capacidades sólidos, fiables y eficaces valiéndose de la IA.

Con sus principios éticos en materia de IA, el Reino Unido pretende fomentar la confianza en estas tecnologías y sus aplicaciones a fin de aprovechar plenamente el potencial de la colaboración entre humanos y máquinas reduciendo al mismo tiempo los riesgos asociados a su utilización, uso abusivo o desuso y evitando consecuencias indeseadas. Con este planteamiento, el Reino Unido pretende aprovechar la innovación y la creatividad de los sectores de la defensa y la industria para posibilitar una utilización ambiciosa de soluciones basadas en la IA.

El Gobierno del Reino Unido tiene claro que cualquier uso de la IA que se haga en el país para mejorar los procesos, los sistemas o las capacidades militares de defensa estará sujeto a la legislación nacional e internacional. Las fuerzas armadas del Reino Unido procuran sistemáticamente cumplir sus obligaciones legales en todas sus actividades, desde el derecho laboral a la privacidad, las adquisiciones y el derecho de los conflictos armados, también conocido como derecho internacional humanitario. Cuentan con prácticas y procesos sólidos para garantizar que sus

actividades y su personal respeten la ley. Estas prácticas y procesos están y seguirán aplicándose a las capacidades de IA. El despliegue de capacidades de IA en conflictos armados debe atenerse plenamente al derecho internacional humanitario y cumplir sus cuatro principios básicos (distinción, necesidad, humanidad y proporcionalidad). Tenemos claro que utilizar cualquier sistema o arma que no cumpla esos principios fundamentales contravendría el derecho internacional.

También son cruciales la responsabilidad y la rendición de cuentas mediante una implicación humana adecuada en función del contexto. Esa implicación humana en función del contexto es necesaria para cumplir nuestras políticas, principios éticos y obligaciones contraídas en virtud del derecho internacional humanitario. La naturaleza de la implicación humana variará en función de la naturaleza de la capacidad, el entorno operacional y el contexto de uso. El Reino Unido velará por que sus armas nucleares estén en todo momento bajo control político humano.

Contribución del Reino Unido a las iniciativas internacionales

La estabilidad mundial requiere un desarrollo ambicioso, pero responsable, de la IA militar. La comprensión por parte de la comunidad internacional de los riesgos, las salvaguardas y las normas relacionadas con el uso de la IA en el contexto militar sigue evolucionando. Esos riesgos son de naturaleza intrínsecamente internacional, por lo que requieren una respuesta global.

El Reino Unido ha estado a la vanguardia de las iniciativas internacionales dirigidas a fomentar un desarrollo y un uso seguros y responsables de la IA. Tiene el orgullo de haber acogido la primera Cumbre sobre Seguridad de la Inteligencia Artificial, en la que se acordó la Declaración de Bletchley sobre la seguridad de la IA, y de haber impulsado la elaboración del Informe Internacional sobre la Seguridad de la IA, primera síntesis exhaustiva de la documentación mundial existente sobre riesgos y capacidades de los sistemas avanzados de IA, que se publicó en febrero de 2025 y aporta ideas cruciales para alimentar el debate internacional, como el aprovechamiento de la IA para la paz y la seguridad. Apoyamos los esfuerzos realizados en el marco del Pacto Digital Global para cerrar las brechas digitales y mejorar la gobernanza internacional de la IA en beneficio de la humanidad.

El Reino Unido apoya activamente las iniciativas internacionales para impulsar actuaciones en el ámbito militar. Hemos apoyado el trabajo de organizaciones como RAND Europe, la Universidad de California en Berkeley y la Comisión Mundial sobre la Inteligencia Artificial Responsable en el Ámbito Militar para reunir a expertos diversos y ampliamente reconocidos con el fin de estudiar estas cuestiones, asimilar las ideas más recientes y ofrecer a los encargados de formular políticas recomendaciones viables con las que avanzar.

El Reino Unido sigue participando activamente en los debates internacionales sobre cuestiones de defensa y seguridad relacionadas con la IA y compartiendo sus experiencias en la concepción y puesta en práctica de modalidades seguras y responsables de aplicación militar de la IA. El Reino Unido se congratula de que, gracias a iniciativas como las Cumbres sobre la Inteligencia Artificial Responsable en el Ámbito Militar, de las que el Reino Unido fue coanfitrión en 2024, y la Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar, propuesta por los Estados Unidos de América, se haya avanzado en la comprensión de las oportunidades y los riesgos estratégicos que plantea la IA y de la manera de abordarlos con medidas adecuadas que promuevan su aplicación segura y responsable. La ética y el control de la IA son cuestiones dinámicas que requieren un compromiso, una colaboración y una iteración continuas.

De cara al futuro

El Reino Unido espera que se aprovechen los avances logrados hasta la fecha respecto de los procesos existentes, entre otras cosas gracias a los debates mantenidos en las Naciones Unidas sobre la base del informe del Secretario General y centrados en medidas concretas. Dada la naturaleza de la IA en el contexto militar, será crucial abordar estas cuestiones de manera inclusiva y multipartita, integrando los conocimientos técnicos, militares y jurídicos de los Estados, el sector privado, el mundo académico y la sociedad civil.

Aunque disponemos de abundante información, nuestra comprensión colectiva de las aplicaciones e implicaciones militares sigue siendo escasa y persisten importantes lagunas y malentendidos sobre la naturaleza y las capacidades de la IA. Es necesario seguir trabajando para desarrollar la capacidad de los Estados, mejorar nuestra comprensión colectiva de las implicaciones y los posibles riesgos y retos estratégicos que plantea la IA en el ámbito militar y concertar unos términos que se acepten universalmente para mantener debates constructivos. Los debates deben centrarse en medidas y prácticas concretas, eficaces y apropiadas que permitan gestionar los riesgos e incluyan, entre otras cuestiones, salvaguardias y normas de comportamiento, nuevos cauces de comunicación y mecanismos de transparencia con los que reducir el riesgo de malas interpretaciones, doctrinas actualizadas, medidas de fomento de la confianza y acuerdos de control de armamentos que reflejen la impronta de la IA militar.

República de Corea

[Original: inglés]
[11 de abril de 2025]

Como tecnología instrumental, la inteligencia artificial (IA) tiene potencial para transformar radicalmente múltiples dimensiones de los asuntos militares, de la toma de decisiones y la recopilación de información a los sistemas de logística, vigilancia y mando y control. Su rápida evolución hace que los Estados tengan cada vez más interés por aprovechar esta tecnología en el ámbito militar.

Las capacidades de la IA y los sistemas que funcionan con ella, cuya integración en las operaciones militares es cada vez mayor, ofrecen oportunidades pero también plantean problemas, sobre todo para la paz y la seguridad internacionales. Estos avances plantean importantes cuestiones humanitarias, jurídicas, de seguridad, tecnológicas y éticas.

Las opiniones expuestas en la presente comunicación se centran específicamente en ámbitos distintos de los sistemas de armas autónomas letales.

Oportunidades de la inteligencia artificial en el ámbito militar

La IA y los sistemas que la integran, como los empleados en actividades de inteligencia, vigilancia y reconocimiento y apoyo para la toma de decisiones, permiten aumentar la conciencia situacional mejorar la precisión y la exactitud y ganar en eficacia gracias al procesamiento de datos a gran escala, la optimización y la generación de conocimientos predictivos, por lo que pueden contribuir a mantener y promover la paz y la seguridad internacionales.

1. Mejorar la aplicación del derecho internacional humanitario y contribuir a la protección de los civiles y los bienes de carácter civil en los conflictos armados

Los sistemas de inteligencia, vigilancia y reconocimiento y apoyo para la toma de decisiones que funcionan con IA pueden mejorar la aplicación de los principios fundamentales del derecho internacional humanitario —distinción, proporcionalidad y precauciones en el ataque— al permitir evaluaciones más precisas del campo de batalla y mejorar la conciencia situacional. La IA puede ayudar a distinguir entre combatientes y no combatientes y a determinar los posibles daños colaterales utilizando información oportuna y bien fundada. Al permitir un mejor conocimiento del campo de batalla, por ejemplo señalando la presencia de civiles, la IA ayuda a determinar la necesidad y la conveniencia de tomar precauciones proteger a los civiles y las infraestructuras civiles.

2. Apoyo a las operaciones de mantenimiento de la paz

La IA puede facilitar el seguimiento de los acuerdos de alto el fuego y los acuerdos de paz y posibilitar mecanismos de alerta temprana con los que detectar posibles infracciones, reforzando así la eficacia y la seguridad de las misiones. La República de Corea ha puesto en marcha en la Unidad Hanbit de la Misión de las Naciones Unidas en Sudán del Sur un proyecto piloto de campamento inteligente para mejorar la seguridad, la eficacia y la capacidad operacional de los campamentos de mantenimiento de la paz de las Naciones Unidas mediante el uso la IA y otras tecnologías emergentes.

3. Mejora de las capacidades de verificación y supervisión de los regímenes de control de armamento y cumplimiento

La IA puede mejorar las capacidades de los mecanismos internacionales de verificación encargados de supervisar el cumplimiento de los acuerdos de control de armamentos y no proliferación. El Organismo Internacional de Energía Atómica puede aprovechar la IA para aumentar la eficacia de los procesos de salvaguardia, en particular los que entrañan clasificar datos, buscar patrones y detectar valores atípicos en los datos. Los sistemas que funcionan con IA también pueden ayudar a detectar signos tempranos de uso de armas químicas o biológicas y descubrir tácticas cada vez más sofisticadas de elusión de las sanciones, reforzando así los regímenes internacionales de no proliferación.

Además de las oportunidades señaladas, la IA puede contribuir a mitigar riesgos estratégicos (como los errores de cálculo, los malentendidos y las escaladas accidentales) al permitir un mejor análisis del comportamiento de los agentes y una mayor capacidad de detección y respuesta proactiva. También puede facilitar el desarrollo de capacidades con las que mejorar la posición en ciberdefensa, proteger infraestructuras nacionales críticas y combatir el terrorismo, entre otras cosas.

Problemas que plantea la inteligencia artificial en el ámbito militar

La aplicación militar de la IA militar podría generar nuevos problemas o agudizar los existentes si no se desarrolla, despliega y utiliza de forma responsable.

Esos problemas pueden surgir de características técnicas y operativas de la IA, como su carácter de caja negra, que hace difícil entender cómo y por qué se obtienen resultados específicos, limitando en consecuencia su inteligibilidad y trazabilidad. Los defectos de concepción y los sesgos inopinados en los datos, los algoritmos o la arquitectura del sistema pueden provocar fallos de funcionamiento o resultados que se desvían de los objetivos previstos. Problemas de dependencia excesiva de los sistemas de IA, como el sesgo de automatización, o de falta de formación pueden dar lugar a

complicaciones asociadas a la falta del discernimiento y la implicación humanos adecuados. Estos factores pueden hacer que aumenten las probabilidades de que se produzcan errores de cálculo, interpretaciones erróneas o escaladas accidentales de conflictos, lo que pondría en riesgo la paz y la seguridad internacionales.

Al ser de doble uso, las tecnologías de IA podrían entrañar mayor riesgo de utilización indebida o abuso por parte de actores irresponsables con malas intenciones. Por ejemplo, en el ámbito cibernetico, podrían intensificarse las campañas de desinformación y los ciberataques realizados con ayuda de IA, como los envenenamientos de datos y las usurpaciones de dirección electrónica. Además, estas tecnologías podrían ser utilizadas por agentes irresponsables para facilitar el desarrollo de nuevas armas químicas o biológicas, lo que generaría problemas de proliferación y amplificaría los riesgos para la paz y la seguridad internacionales.

Aplicación de la inteligencia artificial responsable en el ámbito militar

Para aprovechar sus ventajas y oportunidades afrontando los riesgos y problemas que llevan consigo, la IA y los sistemas que funcionan con ella en el ámbito militar deben desarrollarse, desplegarse y utilizarse de forma responsable a lo largo de todo su ciclo de vida.

La República de Corea está decidida a garantizar y promover la aplicación responsable de la IA en el ámbito militar mediante principios y medidas clave, como los siguientes:

- La IA debe ser ética y estar centrada en el ser humano.
- Las capacidades de IA en el ámbito militar deben utilizarse con arreglo al derecho internacional vigente, incluido el derecho internacional humanitario y de los derechos humanos.
- Los humanos siguen siendo responsables del uso y los efectos de las aplicaciones de la IA en el ámbito militar; la responsabilidad y la rendición de cuentas en ningún caso podrán delegarse en las máquinas.
- Debe garantizarse la fiabilidad y la credibilidad de las aplicaciones de la IA estableciendo las debidas salvaguardias para reducir el riesgo de mal funcionamiento y consecuencias no deseadas, como las derivadas de los sesgos en los datos, en los algoritmos y en otros aspectos.
- Es necesario mantener la intervención humana adecuada en el desarrollo, el despliegue y el uso de la IA en el ámbito militar con medidas adecuadas de raciocinio y control humanos sobre el uso de la fuerza, entre otras.
- El personal competente debe poder comprender, explicar y rastrear adecuadamente los resultados generados por la IA y los sistemas que funcionan con ella en el ámbito militar y de confiar en ellos. Hay que seguir trabajando para mejorar la explicabilidad y la trazabilidad de la IA en el ámbito militar.

La República de Corea patrocina debates y diálogos sobre el desarrollo de medidas que garanticen una IA responsable en el ámbito militar, por ejemplo marcos normativos internacionales; protocolos rigurosos de prueba y evaluación; procesos exhaustivos de verificación, validación y acreditación; mecanismos robustos de supervisión nacional; procesos de seguimiento constante; programas y actividades de capacitación completos; refuerzo de la ciberseguridad; y marcos claros de rendición de cuentas.

Establecer sólidas medidas de control y seguridad es crucial para evitar que agentes irresponsables adquieran y utilicen indebidamente capacidades de IA o sistemas basados en esta tecnología que podrían resultar peligrosos en el ámbito militar.

La República de Corea alienta la concepción de medidas eficaces de fomento de la confianza y medidas adecuadas de reducción del riesgo, así como el intercambio de información y las consultas sobre buenas prácticas y lecciones aprendidas entre los Estados.

La República de Corea subraya la necesidad de evitar que haya agentes estatales y no estatales que utilicen capacidades de IA para contribuir a la proliferación de armas de destrucción masiva y hace hincapié en que esas capacidades no deben entorpecer las iniciativas de control de armamentos, desarme y no proliferación. Es crucial mantener el control y la implicación humanos en todas las acciones críticas para informar y ejecutar las decisiones soberanas relativas al empleo de armas nucleares, sin perjuicio del objetivo último de un mundo libre de armas nucleares.

Las capacidades de IA y los sistemas que utilizan esta tecnología en el ámbito militar deben desarrollarse, desplegarse y utilizarse de modo que mantengan y no socaven la paz y la seguridad internacionales.

Futura gobernanza de la inteligencia artificial en el ámbito militar

Para planear la futura gobernanza de la IA en el ámbito militar, es crucial promover una comprensión común de esta tecnología (sus capacidades y limitaciones), de sus posibles aplicaciones en el ámbito militar y de sus implicaciones para la paz y la seguridad internacionales.

También es importante crear capacidades, especialmente en los países en desarrollo, a fin de promover su plena participación en los debates sobre gobernanza y facilitar un planteamiento responsable y una idea común del desarrollo, el despliegue y el uso de la IA en el ámbito militar. El intercambio de conocimientos, buenas prácticas y lecciones aprendidas puede asimismo facilitar un entendimiento común.

La IA está evolucionando con rapidez y los mecanismos de gobernanza deben ser lo suficientemente flexibles para adaptarse a esa evolución. La República de Corea preconiza también un abordaje equilibrado que tenga en cuenta tanto las oportunidades como los riesgos. Plantear esta cuestión centrándose excesivamente en los riesgos o en términos restrictivos puede asfixiar la innovación y disipar posibles contribuciones de la IA para la paz y la seguridad internacionales. Antes que entorpecer la innovación, la gobernanza futura debería más bien promoverla y facilitar la aplicación responsable de la IA en el ámbito militar.

Habida cuenta de que la comunidad internacional apenas está empezando a entender las implicaciones que el uso de la IA en el ámbito militar tiene para la paz y la seguridad internacionales y en vista del estado actual del desarrollo tecnológico y normativo, sería prematuro definir de forma restrictiva el curso que seguirá la gobernanza de la IA o establecer instrumentos o normas jurídicamente vinculantes sin una comprensión común y compartida de lo que sería una IA responsable en el ámbito militar. La República de Corea cree que los debates sobre este tema deben ser realistas y discurrir paulatinamente, guiándose por un diálogo constante.

Reconociendo que la innovación en IA está siendo impulsada por el sector privado, la República de Corea cree que las futuras labores de gobernanza deben adoptar un enfoque abierto e inclusivo que integre la participación de múltiples interesados, como las empresas del sector, el mundo académico, la sociedad civil y las organizaciones regionales e internacionales.

La República de Corea valora positivamente las iniciativas nacionales, regionales y mundiales desplegadas para abordar las oportunidades y retos que acarrea la IA en el ámbito militar, como la elaboración de estrategias, disposiciones

legislativas, principios, normas, políticas y medidas nacionales sobre la cuestión, y reconoce la importancia de promover el diálogo a todos los niveles.

Para garantizar la aplicación responsable de la IA en el ámbito militar, la República de Corea creó, en 2022 y 2025 respectivamente, la División de Políticas de Datos y el equipo de Políticas de IA de Defensa dentro del Ministerio de Defensa Nacional. En 2024, el Ministerio puso en marcha el Comité de Datos de Defensa e IA, que es el órgano deliberativo y decisivo de más alto nivel.

A fin de promover el diálogo, la República de Corea organizó, junto con los Países Bajos, Singapur, Kenya y el Reino Unido, la Segunda Cumbre sobre la Inteligencia Artificial Responsable en el Ámbito Militar, que tuvo lugar en Seúl en septiembre de 2024. Estas Cumbres y una serie de consultas regionales sobre el mismo tema celebradas en 2024 sirvieron de plataforma para fomentar el intercambio de conocimientos técnicos, el diálogo inclusivo y el entendimiento mutuo. Próximamente, la Tercera Cumbre sobre la Inteligencia Artificial Responsable en el Ámbito Militar, que se celebrará en España en septiembre de 2025, y las consultas regionales sobre el mismo tema que se realizarán ese mismo año, seguirán guiando los esfuerzos de la comunidad internacional en pos de una aplicación responsable de la IA en el ámbito militar.

La República de Corea considera que los debates sobre la aplicación responsable de la IA en el ámbito militar que mantienen órganos de las Naciones Unidas, como la Primera Comisión de la Asamblea General y la Comisión de Desarme, deben complementar otras iniciativas sobre el particular ajenas a la Organización, como el proceso de Cumbres sobre la Inteligencia Artificial Responsable en el Ámbito Militar, la Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar y el Grupo de Expertos Gubernamentales sobre las Tecnologías Emergentes en el Ámbito de los Sistemas de Armas Autónomos Letales, que, a su parecer, se refuerzan y complementan mutuamente.

La gobernanza de los datos también es crucial. Los datos son esenciales para las actividades de capacitación, despliegue y evaluación relativas a los sistemas de IA, por lo que los interesados han de seguir estudiando mecanismos adecuados de gobernanza que tengan normas y procedimientos claros de recopilación, almacenamiento, tratamiento, intercambio y eliminación de los datos, así como dispositivos para su protección.

Serbia

[Original: inglés]
[4 de abril de 2025]

El desarrollo y la aplicación de la inteligencia artificial constituyen un importante factor de cambio en la forma en que se llevan a cabo las operaciones militares en el mundo actual. Ofrece nuevas posibilidades, pero también plantea nuevos retos para la estabilidad internacional y la paz y la seguridad en el ámbito militar, por lo que resulta necesario iniciar la creación de un marco internacional adecuado para regular su aplicación.

1. Posibilidades y ventajas del uso de la inteligencia artificial en el ámbito militar

La aplicación de la inteligencia artificial en el contexto militar no letal puede mejorar muchos aspectos de las operaciones militares:

- a) Aumentando el grado de conciencia operacional;
- b) Mejorando la calidad y la rapidez del proceso de toma de decisiones;

- c) Mejorando la calidad de los datos de inteligencia y reconocimiento mediante el procesamiento rápido de datos y permitir una detección rápida de las amenazas;
- d) Respaldando la protección de civiles y no combatientes en conflictos militares;
- e) Apoyando las operaciones y misiones de paz mediante la supervisión de los acuerdos de alto el fuego y el pronóstico de la dinámica de los conflictos;
- f) Mejorando los procesos y procedimientos de mantenimiento predictivo y optimización logística reduciendo costos y ahorrando recursos.

2. Principales retos y amenazas del uso de la inteligencia artificial en el ámbito militar

El desarrollo y la integración de la inteligencia artificial en los sistemas de combate y no combate plantean importantes retos para la paz y la estabilidad internacionales y para el derecho internacional humanitario, principalmente los siguientes:

- a) Riesgos técnicos y fallos de funcionamiento debidos a errores de aplicación en entornos dinámicos, que pueden poner en peligro vidas humanas, causar daños materiales y socavar la aplicación del derecho internacional humanitario;
- b) Riesgos jurídicos y éticos asociados al cumplimiento del derecho internacional, especialmente en lo que respecta a la aplicación de sus principios, como la distinción, la proporcionalidad y las precauciones en la selección de objetivos;
- c) Falta de normas explícitas de atribución de responsabilidad en actuaciones y actividades realizadas con inteligencia artificial;
- d) La imperfección de los algoritmos puede dar lugar a sesgos, errores en la toma de decisiones y resultados discriminatorios, ya que aplicar series de datos no representativos puede llevar a identificaciones erróneas de civiles o a amenazas para grupos étnicos o nacionales;
- e) La aplicación de los algoritmos de la inteligencia artificial puede suscitar la impresión equivocada de que la responsabilidad de las personas que participan en operaciones es menor;
- f) Riesgos estratégicos de las decisiones adoptadas por medio de la inteligencia artificial a partir de premisas incorrectas;
- g) Convergencia e integración no selectivas con las nuevas tecnologías, en particular en los ámbitos de la información y las ciberoperaciones o la aplicación de medios nucleares, químicos y biológicos;
- h) Falta de profesionales que desarrollem, organicen y apliquen de manera responsable sistemas de inteligencia artificial en los conflictos;
- i) Uso indebido de la inteligencia artificial en operaciones de información para generar y difundir desinformación, lo que puede instigar conflictos y agravar tensiones.

3. Creación de un marco jurídico y ético

Habida cuenta los riesgos y desafíos señalados, es necesario establecer marcos jurídicos y éticos vinculantes en el seno de la comunidad internacional a fin de:

- a) Promover y tratar de establecer en el seno de las Naciones Unidas un diálogo que permita reforzar el cumplimiento de las normas del derecho internacional

humanitario, lo que entraña establecer normas, reglas y principios jurídicos internacionales que garanticen que el desarrollo y la aplicación de sistemas de inteligencia artificial se ajusten a los principios del derecho internacional humanitario (distinción, proporcionalidad y precauciones para proteger a las personas que no participan en conflictos armados);

b) Iniciar un proceso de análisis de la legalidad en la aplicación de sistemas y armamentos respecto de las modalidades aprobadas de aplicación de la inteligencia artificial;

c) Garantizar la protección de la vida y la libertad de las personas durante los conflictos armados y su privacidad en tiempos de paz, en particular en situaciones de seguimiento;

d) Reforzar los mecanismos de las Naciones Unidas introduciendo la obligación de analizar el riesgo de aplicar la inteligencia artificial con fines militares, modernizando la Conferencia de Desarme, armonizando el trabajo de la Comisión de Desarme, estableciendo nuevos órganos especializados de las Naciones Unidas y ampliando las iniciativas para un uso responsable de la inteligencia artificial ya en marcha en las Naciones Unidas;

e) Entablar en las Naciones Unidas un diálogo que permita definir el uso responsable de la inteligencia artificial en el ámbito militar y establecer protocolos de seguridad para su aplicación (ensayo, evaluación, validación y verificación);

f) Elaborar medidas para adecuar la intervención del sector privado con los principios del derecho internacional humanitario durante el desarrollo, la implantación y la aplicación de sistemas y servicios de inteligencia artificial en ámbitos militares;

g) Ampliar los institutos y documentos existentes de las Naciones Unidas sobre recomendaciones relativas a la ética del desarrollo y la aplicación de la inteligencia artificial introduciendo directrices específicas sobre la conducción de conflictos.

La aplicación de sistemas de inteligencia artificial en el contexto de conflictos armados internacionales exige una amplia acción multilateral de la comunidad internacional para promover la responsabilidad en su uso. Las Naciones Unidas deberían liderar las labores de promoción del diálogo, las normas y el desarrollo de capacidades de la comunidad internacional para evitar la fragmentación y lograr una gestión adecuada.

Singapur

[Original: inglés]
[11 de abril de 2025]

Como Estado pequeño, Singapur siempre ha apoyado el sistema multilateral basado en normas y el papel de las Naciones Unidas. Las Naciones Unidas sientan las bases del derecho y las normas internacionales. Las instituciones, los sistemas y las leyes multilaterales son vitales para la supervivencia de todos los Estados, en particular de los pequeños.

Singapur considera que las capacidades de inteligencia artificial (IA) en el ámbito militar y los sistemas que funcionan con esta tecnología deben aplicarse durante todo su ciclo de vida de forma responsable y de conformidad con el derecho internacional aplicable, en particular, el derecho internacional humanitario.

La IA podría ser beneficiosa en el ámbito militar al mejorar la precisión y la conciencia situacional, lo que permitiría reducir los daños colaterales en civiles y bienes de carácter civil. Sin embargo, a falta de unos marcos de gobernanza adecuados, también puede plantear el riesgo de intensificar conflictos y errores de cálculo. A Singapur le parece importante que la comunidad internacional se ocupe de esta cuestión.

Postura de Singapur sobre la gobernanza de la inteligencia artificial en el ámbito militar

Uno de los principales objetivos de la Estrategia Nacional de IA 2.0 de Singapur es fomentar un entorno de confianza que proteja a los usuarios y facilite la innovación. Con ese fin, varios sectores gubernamentales, incluido el de la defensa, están elaborando marcos de gobernanza que permitan aprovechar los beneficios de la IA reduciendo al mismo tiempo los problemas que pueda entrañar su uso.

Mediante consultas con expertos en tecnología de defensa, responsables de planificación militar, expertos en derecho internacional y profesionales de la política, Singapur elaboró unos principios nacionales sobre inteligencia artificial en el ámbito militar, que se presentaron en 2021 y abordan cuatro ámbitos clave de preocupación:

- a) **Responsabilidad.** En primer lugar, hay que abordar los riesgos de comportamiento de la IA emergente. Los sistemas de IA deben tener usos previstos bien definidos y tanto los desarrolladores como los usuarios son responsables de sus resultados;
- b) **Fiabilidad.** En segundo lugar, hay que abordar el riesgo de errores o imprecisiones en los resultados de los sistemas de IA, que han ponerse a prueba a fin de garantizar debidamente su uso previsto y concebirse de modo que se minimicen los sesgos involuntarios y se generen resultados coherentes;
- c) **Solidez.** En tercer lugar, deben abordarse los riesgos asociados al uso de la IA por agentes malintencionados. Al concebir sistemas de IA han de tenerse presentes las amenazas ciberneticas y de fuerzas en conflicto. El proceso de desarrollo debe estar bien documentado para facilitar la explicabilidad y contrarrestar el efecto “caja negra”;
- d) **Seguridad.** En cuarto lugar, debemos centrarnos en el riesgo de fallo de la IA en contextos críticos para la seguridad. Los sistemas de IA deben ser seguros, no solo para las plataformas desplegadas, sino también para los activos y el personal circundantes.

Estos principios rectores fundamentan la postura de Singapur sobre el desarrollo, las pruebas, la formación y el despliegue de sistemas que utilizan la IA con fines militares.

Iniciativas internacionales y regionales sobre inteligencia artificial en el ámbito militar

Singapur ha participado activamente en iniciativas internacionales sobre la gobernanza de la IA en el ámbito militar. En 2023, suscribió el Llamamiento a la Acción sobre Inteligencia Artificial Responsable en el Ámbito Militar y la Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar. En 2024, copatrocinó la Cumbre sobre la Inteligencia Artificial Responsable en el Ámbito Militar, celebrada en Seúl (República de Corea), donde se aprobó el Plan de Acción sobre Inteligencia Artificial Responsable en el Ámbito Militar.

Singapur también reconoce que las iniciativas regionales son importantes para que los debates sobre la IA en el ámbito militar sean inclusivos y tengan en cuenta circunstancias específicas. Singapur copatrocinó las consultas regionales sobre la IA responsable en el ámbito militar que se celebraron en 2024 para Asia y sirvieron de foro a los países de la región para intercambiar opiniones sobre las oportunidades y los riesgos que plantea la IA en el ámbito militar y otros temas.

En febrero de 2025, durante la Estancia de Encuentro de Ministros de Defensa de la Asociación de Naciones de Asia Sudoriental (ASEAN) que tuvo lugar en Penang (Malasia), Singapur y otros Estados miembros de la ASEAN aprobaron una declaración conjunta sobre cooperación para la aplicación de la IA en el sector de la defensa en la que se comprometieron a promover el uso responsable de la IA y la rendición de cuentas al respecto, a intercambiar información para profundizar en la comprensión y la concienciación regionales sobre las implicaciones de la IA en este sector y a compartir mejores prácticas y lecciones aprendidas entre Estados miembros de la ASEAN.

Futuros debates sobre inteligencia artificial y paz y seguridad internacionales en las Naciones Unidas

Singapur considera que los debates que se organicen ulteriormente sobre la base del apoyo de la comunidad internacional a esta resolución han de ser abiertos e inclusivos, por lo que vería con buena disposición que, en el seno de las Naciones Unidas, se crease un grupo de trabajo de composición abierta dedicado a la IA en el ámbito militar. De crearse, este grupo de trabajo debería adoptar un enfoque multipartito que permitiera contar con la participación de expertos en tecnología, responsables de planificación militar, juristas internacionales y profesionales de la política, entre otros. Reafirmamos nuestra voluntad de trabajar con todos los Estados Miembros para avanzar en la aplicación responsable de la IA en el ámbito militar.

Suiza

[Original: inglés]
[11 de abril de 2025]

1. Oportunidades y riesgos

Es probable que la inteligencia artificial (IA) transforme muchos aspectos de los asuntos militares. Se augura que aportará a las tareas y operaciones militares mayor fiabilidad, eficacia, precisión, seguridad y solidez, particularmente en los ámbitos de la conciencia situacional, la toma de decisiones, la inteligencia, la vigilancia y reconocimiento, la logística y las cadenas de suministro, las actividades de adiestramiento y simulación y las tareas de mando y control, gracias al análisis de grandes conjuntos de datos y a la posibilidad de tomar decisiones más rápidas y fundamentadas. Por ejemplo, en el ámbito de la vigilancia y el reconocimiento, la IA puede analizar imágenes de drones y satélites para detectar movimientos más rápido que los analistas humanos. También podría mejorar el reconocimiento de objetivos procesando datos de sensores para distinguir entre fuerzas amigas y hostiles. En el ámbito de la logística, la IA puede optimizar las cadenas de suministro, predecir fallos en los equipos y garantizar que los recursos lleguen al lugar adecuado en el momento oportuno. En materia de apoyo para la toma de decisiones, las simulaciones de IA pueden proporcionar a los mandos pronósticos y resultados potenciales para orientar la planificación estratégica. Los sistemas de adiestramiento y simulación que funcionan con IA ofrecen entornos realistas y adaptables que permiten preparar mejor a los soldados. Por último, la IA puede respaldar las tareas de mando y control

agilizando el flujo de información, mejorando la toma de decisiones y potenciando la coordinación entre unidades. La IA también puede ayudar a la detección de amenazas, la ciberseguridad, el mantenimiento de la paz, la verificación del control de armamentos y la distensión de conflictos mediante sistemas de alerta temprana, análisis predictivos y mecanismos de supervisión, contribuyendo así a la estabilidad y la seguridad. Con todo, aunque estos avances pueden aportar beneficios a las fuerzas armadas, la integración de la IA en el ámbito militar también plantea importantes preocupaciones y posibles riesgos.

Cuando se utiliza de forma responsable en los conflictos armados, la IA puede contribuir a reforzar el cumplimiento del derecho internacional humanitario y a fortalecer la protección de los civiles y los bienes de carácter civil, por ejemplo al mejorar las evaluaciones de riesgos o incrementar la precisión de los objetivos para reducir daños colaterales. Sin embargo, hay aplicaciones militares de la IA en conflictos armados, especialmente las que entrañan alto riesgo, que plantean también graves problemas jurídicos, humanitarios, éticos, de seguridad y de estabilidad estratégica que deben abordarse, como los siguientes:

- **Errores de selección de objetivos.** Aunque, técnicamente, la IA puede identificar objetos o personas basándose en sus datos de adiestramiento, la comprensión del contexto y los juicios de valor necesarios para cumplir el derecho internacional plantean una dificultad particular, lo que puede conducir a identificaciones erróneas de objetos y personas como blancos militares y, por tanto, a ataques ilícitos o involuntarios.
- **Riesgos de escalada.** En crisis de rápida evolución, una herramienta de apoyo para la toma de decisiones con funcionamiento de caja negra podría recomendar una acción de agresión sin un fundamento claro. Al no tener capacidad de explicación, los mandos podrían seguir ciegamente una orientación errónea o perder un tiempo crítico cuestionándola.
- **Interpretación errónea de intenciones.** Un sistema de IA que evalúe los riesgos asociados a acciones humanas o a objetos puede plantear problemas (legales y de seguridad), especialmente cuando las evaluaciones se basan en patrones derivados de comportamientos y circunstancias pasados y no se someten a control y juicio humanos en función del contexto. Por ejemplo, un sistema de IA que esté siguiendo el comportamiento del adversario puede clasificar erróneamente movimientos ordinarios de contingentes como hostiles, debido a datos viciados, lo que podría desencadenar acciones preventivas e intensificar involuntariamente las tensiones.

Estos riesgos ponen de relieve la obligación de garantizar que se cumpla el derecho internacional vigente, en particular el derecho internacional humanitario, pero también la necesidad imperiosa de seguir dialogando y estudiando la cuestión para comprender mejor los riesgos y desafíos y las medidas que podrían resultar necesarias, así como para analizar la necesidad, el valor añadido y la viabilidad de desarrollar estructuras adicionales de gobernanza normativa, por ejemplo promulgando leyes nacionales, mejorando prácticas, normas, criterios o instrumentos internacionales o estableciendo directrices operacionales.

2. Marco jurídico

Como cualquier otra tecnología, la IA no se desarrolla y aplica en un vacío legal. En el ámbito militar, su desarrollo, despliegue y utilización debe atenerse plenamente al derecho internacional vigente, en particular la Carta de las Naciones Unidas, el derecho internacional humanitario y de los derechos humanos y demás marcos jurídicos pertinentes. Ninguna tecnología ha de poner en cuestión en ningún caso la

validez del derecho internacional. El derecho internacional vigente, en particular la Carta de las Naciones Unidas en su totalidad, el derecho internacional de los derechos humanos y el derecho internacional humanitario, ha de cumplirse y respetarse.

Los Estados y las partes en conflicto deben respetar y hacer respetar el derecho internacional humanitario en toda circunstancia, también cuando utilicen IA en operaciones militares. Por ello, las aplicaciones de la IA en el ámbito militar deben concebirse de forma que refuercen el cumplimiento del derecho internacional humanitario y la protección de los civiles y los bienes de carácter civil. Una forma de lograrlo sería garantizar que los sistemas de IA primen la precisión, la minimización de los daños y la responsabilidad, por ejemplo, mediante procesos rigurosos de selección, validación y verificación de objetivos. Además, el uso de IA debe reforzar el cumplimiento de la obligación legal de que en las operaciones militares se tomen todas las precauciones posibles a fin de evitar o, al menos, minimizar los daños incidentales, ayudando a los mandos a proteger a los civiles y bienes de carácter civil mientras duran las hostilidades, por ejemplo, mejorando las evaluaciones de riesgos.

Una medida esencial es adiestrar las aplicaciones militares de la IA con series de datos que permitan utilizarlas cumpliendo plenamente el derecho internacional. Más allá de la conducción de hostilidades, las aplicaciones militares de la IA deben cumplir todas las normas y principios pertinentes del derecho internacional humanitario al realizar cualquier otra tarea que esté sujeta a sus disposiciones, como la detención y el internamiento de personas o el control de masas y la aplicación de medidas de seguridad pública en territorios ocupados.

Al desarrollar y utilizar la IA en el ámbito militar, se corre el riesgo de incorporar en el diseño de los sistemas o en los datos de adiestramiento interpretaciones jurídicas excesivamente laxas (por ejemplo, ampliando la definición de los objetivos lícitos o incrementando el umbral de los daños incidentales aceptables). Al aplicarse a mayor escala, tales interpretaciones podrían minar progresivamente el objetivo de protección del derecho internacional humanitario y provocar un aumento sustancial de los daños a civiles. Este riesgo pone de relieve la importancia de salvaguardar la integridad de las normas jurídicas, que debe seguir siendo una consideración central en la gobernanza, la concepción y el despliegue de aplicaciones militares de IA en el futuro.

3. Premisas y principios

Partiendo del marco jurídico que acaba de esbozarse y basándose en él, y también teniendo en cuenta las preocupaciones humanitarias, éticas, de seguridad y de estabilidad estratégica, deberían profundizarse las premisas y principios siguientes:

1. *Responsabilidad, rendición de cuentas e implicación humanas*

- **Responsabilidad y rendición de cuentas.** Los Estados deben garantizar que la responsabilidad y la rendición de cuentas por las decisiones militares adoptadas con ayuda de IA se mantengan en todo momento en manos humanas, conforme al derecho internacional vigente.
- **Control y juicio humanos en función del contexto.** Las decisiones militares críticas —del centro de mando al campo de batalla—, sobre todo cuando entrañan el uso de la fuerza, deben someterse en todo momento a un control y juicio humanos que tengan en cuenta las circunstancias. En el ámbito militar, la IA puede ayudar a tomar decisiones, pero no debe suplir las consideraciones y valoraciones jurídicas y éticas inherentes a la autonomía cognitiva de decisión. Los Estados únicamente deben integrar estos sistemas en cadenas de mando y

control en las que los humanos sigan ejerciendo un arbitraje y unos niveles de control adecuados. Los sesgos involuntarios deben corregirse en la medida de lo posible.

2. *Fiabilidad, previsibilidad/explicabilidad, solidez*

- **Fiabilidad.** Las aplicaciones militares de la IA deben ser fiable para evitar consecuencias indeseadas o fallos de funcionamiento, sobre todo si pueden perjudicar o dañar a civiles y bienes de carácter civil, y solo han de utilizarse cuando sus efectos y consecuencias puedan preverse en medida razonable.
- **Previsibilidad/inteligibilidad.** Los procesos de decisión asistidos por IA deben ser predecibles y explicables para los responsables de desplegarlos, permitiéndoles comprender y anticipar los comportamientos del sistema.
- **Solidez.** En el ámbito militar, la IA también debe tener solidez –técnica y operativa– para poder desplegarla y utilizarla manteniendo la seguridad.

3. *Reducción del riesgo*

- **Mejorar la conciencia situacional.** La IA debe utilizarse para mejorar el conocimiento del campo de batalla y poder detectar la presencia de civiles a fin de reducir el riesgo de daños.
- **Análisis predictivo.** Los modelos predictivos elaborados con IA deben utilizarse, entre otras cosas, para ayudar a evaluar los riesgos, elaborar estrategias de distensión de conflictos y evitar bajas civiles.
- **Salvaguardas integradas.** La IA en el ámbito militar debe incorporar salvaguardas que minimicen los daños y permitan una intervención humana adecuada en caso de fallos del sistema.

4. *Evitar nuevos motivos de escalada*

- **Estabilidad.** Las aplicaciones militares de la IA solo deben concebirse, desplegarse y utilizarse de forma que no agudicen las tensiones internacionales ni generen nuevos motivos de escalada.
- **Control de armamentos.** La IA, que puede servir para el control de armamentos, no debe socavar las normas e instrumentos existentes de no proliferación, control de armamentos y desarme, ni entorpecer el cumplimiento de normas de este tipo, en particular las relativas a las armas biológicas y nucleares.
- **Gestión de crisis.** En el ámbito militar, la IA podría facilitar la distensión y la gestión de crisis.

5. *Gestión del ciclo de vida de los sistemas militares de inteligencia artificial*

El uso militar responsable de la IA requiere un enfoque global y sensible a los riesgos que aborde los sistemas militares que funcionan con esta tecnología en todas las fases de su ciclo de vida (concepción, desarrollo, pruebas, despliegue, utilización, renovación y desmantelamiento). En cada una de ellas han de tenerse en cuenta sistemáticamente las consideraciones jurídicas, humanitarias, operacionales y técnicas pertinentes. Este enfoque basado en el ciclo de vida es especialmente importante en aplicaciones militares de alto riesgo, como las armas autónomas, la selección de objetivos o el apoyo para la toma de decisiones que podrían acarrear riesgos de lesiones o muertes humanas o daños materiales y, en general, de decisiones sujetas al derecho internacional humanitario. El ciclo de vida de los sistemas en los

que el riesgo es menor, como las herramientas de apoyo administrativo o los sistemas de planificación logística, debe gestionarse evaluando los riesgos específicos de cada contexto.

- Durante la fase de concepción y desarrollo, los Estados deben asegurarse de que los sistemas se adiestren con series de datos representativos y de alta calidad —con un mínimo de sesgos— para que su uso se ajuste plenamente al derecho, las normas y los criterios internacionales y para reducir al mínimo los sesgos no deseados.
- En la fase de prueba y evaluación, deben aplicarse procedimientos rigurosos de validación y verificación que permitan confirmar su fiabilidad, conformidad con la ley y solidez operacional en condiciones realistas.
- En la fase de despliegue y utilización, deben existir salvaguardias que permitan supervisar el funcionamiento de los sistemas, garantizar un control y un juicio humanos adecuados en función del contexto y la intervención humana necesaria.
- En las fases de renovación y adiestramiento, los Estados deben establecer protocolos rigurosos de modificación de los sistemas que entrañen procesos de control de las versiones, revalidación y aprobación formal.
- Para la fase de retirada o desmantelamiento, deben establecerse medidas de desactivación o archivado seguros de los sistemas que eviten su uso indebido, su activación involuntaria o su reutilización.

4. Gobernanza internacional

Suiza destaca la importancia de que las Naciones Unidas acometan un proceso inclusivo y sostenido para consolidar una visión común de los beneficios, los riesgos y problemas de la IA en el ámbito militar y establecer unos principios de uso responsable de esta tecnología. Este proceso debe pues incluir a todos los Estados Miembros y partes interesadas pertinentes, así como a científicos y representantes del sector tecnológico, la sociedad civil y el mundo académico, para garantizar la legitimidad, la competencia y un amplio apoyo. Los procesos de las Naciones Unidas en este ámbito deben ser transparentes, realizarse con regularidad y converger con otras iniciativas pertinentes.

Todas las iniciativas de gobernanza internacional para un uso responsable de la IA en el ámbito militar deben perseguir el propósito general de garantizar el cumplimiento del derecho internacional, en particular del derecho internacional humanitario. También deben tener entre sus prioridades las preocupaciones humanitarias y éticas, el mantenimiento de la estabilidad y la reducción de los riesgos de seguridad. Unos marcos de gobernanza eficaces, unas normas compartidas y un diálogo multilateral sostenido deberían contribuir a evitar escaladas indeseadas, a fomentar la transparencia y la confianza mutua y a reforzar el papel del derecho internacional en tiempos de revolución tecnológica. Al asentar en estos principios la gobernanza de la IA en el ámbito militar, los Estados contribuyen a un entorno de seguridad más predecible, resiliente y apacible.

Podrían tomarse medidas concretas como las siguientes:

- Promover concepciones, definiciones y términos comunes y una visión compartida del alcance de las aplicaciones militares de la IA
- Conocer y entender mejor las oportunidades y los problemas humanitarios, jurídicos, de seguridad y éticos
- Estudiar medidas de transparencia y fomento de la confianza

- Establecer principios, normas, mejores prácticas y otras recomendaciones
- Orientar sobre su aplicación

Ucrania

[Original: inglés]
[11 de abril de 2025]

Ucrania, que ha estado trabajando en el desarrollo y la aplicación de la inteligencia artificial (IA) en diversos ámbitos, como el militar, entiende claramente tanto el potencial que encierra esta tecnología para mejorar el bienestar humano y las capacidades militares, como los importantes riesgos de su uso indebido en la esfera civil y, sobre todo, en la militar. Estos riesgos son especialmente acuciantes en el contexto de la invasión total, no provocada e injustificada, de Ucrania por la Federación de Rusia, que ha violado sistemáticamente las leyes y costumbres de la guerra y el derecho internacional humanitario.

Las iniciativas internacionales encaminadas a alcanzar un consenso mundial sobre el desarrollo, el despliegue y el uso responsables de la IA en los ámbitos civil y militar cuentan con el apoyo y la participación de Ucrania.

Entre las actuaciones que ha llevado a cabo hasta la fecha, Ucrania suscribió en 2023 la Declaración de Bletchley; es uno de los Estados que aprobaron la Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar, que se presentó en la Cumbre sobre la Inteligencia Artificial Responsable en el Ámbito Militar celebrada en La Haya en 2023; secundó el Llamamiento a la Acción por una Inteligencia Artificial Responsable en el Ámbito Militar acordado en la Cumbre sobre la Inteligencia Artificial Responsable en el Ámbito Militar de 2023 y el Plan de Acción sobre Inteligencia Artificial Responsable en el Ámbito Militar, que se aprobó como documento final en la edición de 2024 de dicha Cumbre; se sumó a la Declaración sobre Inteligencia Artificial Inclusiva y Sostenible para las Personas y el Planeta dimanada de la Cumbre de Acción sobre Inteligencia Artificial celebrada en 2025 en París; y copatrocinó las tres resoluciones de la Asamblea General sobre IA aprobadas hasta la fecha, incluida la resolución [79/239](#) sobre la inteligencia artificial en el ámbito militar y sus implicaciones para la paz y la seguridad internacionales.

Ucrania está dispuesta a participar activamente en las nuevas iniciativas mundiales para fomentar el desarrollo seguro, ético y responsable de la IA. También apoya los debates sobre los diferentes aspectos de la IA que celebran los organismos del sistema de las Naciones Unidas, incluido el Consejo de Seguridad.

Ucrania, una nación amante de la paz que no hace reclamaciones territoriales a otras naciones y que, al mismo tiempo, es víctima de una agresión militar rusa y no reconoce reclamaciones territoriales ajenas, desarrolla y utiliza la IA militar con el único fin de reforzar sus capacidades de defensa en ejercicio del derecho de legítima defensa consagrado en la Carta de las Naciones Unidas.

En su uso militar de la IA, Ucrania constata los siguientes grandes riesgos para la paz y la seguridad internacionales:

- La competencia en la integración de la IA en los sistemas de combate y armamento podría desencadenar una carrera armamentista mundial nueva y aún más peligrosa en detrimento de la consecución de los Objetivos de Desarrollo Sostenible y, más concretamente, la aparición de sistemas de armas totalmente autónomos capaces de funcionar sin intervención humana.

- Como ocurre con otras tecnologías digitales, la creciente amenaza que plantean los ciberataques y el aumento de la complejidad y los ámbitos de aplicación de esta tecnología hacen que los sistemas militares que la utilizan sean cada vez más vulnerables a interferencias cibernéticas y manipulaciones de partes interesadas a fin de despojarlas de las características de aplicación y las funciones de uso selectivo previstas.
- Una dependencia excesiva de la IA en la toma de decisiones podría entrañar la pérdida de control humano sobre procesos militares críticos.
- La integración precipitada en los sistemas de armas de herramientas de IA insuficientemente desarrolladas, sobre todo cuando tienen unas capacidades deficientes de identificación de objetivos, puede causar daños indiscriminados y hacer que aumenten las bajas civiles.
- Actualmente no existe un marco multilateral para controlar la proliferación de armas que integran IA.
- Utilizar armamentos con IA integrada sin atenerse a las leyes y costumbres de la guerra y al derecho internacional humanitario plantea graves problemas jurídicos y éticos.

B. Unión Europea

[Original: inglés]
[11 de abril de 2025]

La Unión Europea celebra tener la ocasión de ofrecer sus perspectivas sobre los retos y las oportunidades para la paz y la seguridad internacionales que plantea la inteligencia artificial (IA) en el ámbito militar, de conformidad con la resolución 79/239, que la Asamblea General aprobó el 24 de diciembre de 2024.

Ante todo, la Unión Europea desea recordar su postura de larga data de que el uso de la IA en el ámbito militar debe atenerse al derecho internacional, en particular la Carta de las Naciones Unidas, el derecho internacional humanitario y el derecho internacional de los derechos humanos.

Asimismo, la Unión Europea desea recordar otra de sus posturas tradicionales, a saber, que el uso de la fuerza han de someterse en todo momento al juicio y el control humanos. En cuestiones de IA en el ámbito militar, la responsabilidad y la rendición de cuentas deben permanecer en manos humanas a fin de garantizar que esta tecnología se aplique de forma responsable.

La Unión Europea reconoce que la aplicación de la IA a los sistemas militares conlleva oportunidades y desafíos. El desarrollo de la IA es tan rápido que en la actualidad no es posible predecir las ventajas o riesgos que presentará.

La Unión Europea se congratula de que las Naciones Unidas sigan ocupándose de esta cuestión y de que los foros internacionales pertinentes mantengan debates al respecto. A este respecto, la Unión Europea valora en particular la continuación del proceso de Inteligencia Artificial Responsable en el Ámbito Militar, que se inició en los Países Bajos en 2023 con la primera Cumbre sobre la Inteligencia Artificial Responsable en el Ámbito Militar, seguida en 2024 por la segunda de estas cumbres, que organizó la República de Corea. La Unión Europea celebra que este proceso siga adelante con la celebración de la tercera edición de la Cumbre, que tendrá lugar en España en 2025, y da las gracias a España por organizarla.

La Unión Europea hace notar que el Llamamiento a la Acción para una Inteligencia Artificial Responsable en el Ámbito Militar, de 2023, y el Plan de Acción

para una Inteligencia Artificial Responsable en el Ámbito Militar, de 2024, recibieron el apoyo de todos sus Estados miembros. La Unión Europea considera que abordar esta cuestión mediante procesos inclusivos multipartitos, como se planteó en la Cumbre sobre la Inteligencia Artificial Responsable en el Ámbito Militar, es un enfoque prometedor. La Unión Europea valora positivamente otras contribuciones recientes, como la Cumbre Internacional sobre Inteligencia Artificial y la Cumbre de Acción sobre Inteligencia Artificial, de la que Francia fue anfitriona los días 10 y 11 de febrero 2025. Valora asimismo el trabajo realizado en el marco de la Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar, que constituye una valiosa contribución al debate internacional general sobre las consecuencias de la IA para la paz y la seguridad internacionales.

La Unión Europea considera que los documentos finales y la Declaración Política sobre Uso Responsable de Inteligencia Artificial y Autonomía en el Ámbito Militar, que suscribieron todos sus Estados miembros, son complementarios y revisten gran importancia en el desarrollo de una reflexión global, de la gobernanza y de soluciones prácticas para el uso militar responsable de la IA.

La Unión Europea reconoce que la aplicación de la IA en el ámbito militar tiene ventajas, sobre todo en lo que se refiere a la rapidez, la magnitud y la precisión de las operaciones militares. La IA puede aportar una ventaja táctica gracias a la gestión y el preprocesamiento de vastos conjuntos de datos procedentes de sistemas de vigilancia y armamento, drones e imágenes satelitales, que pueden permitir a los operadores humanos tomar decisiones más rápidas y mejores. Con las aplicaciones de la IA pueden reducirse los costos mejorando la logística o el mantenimiento de los equipos mediante la gestión predictiva. La IA también puede facilitar operaciones militares a gran distancia y mejorar su precisión en entornos inestables.

Pero esas mismas ventajas de las aplicaciones militares de la IA en lo tocante a la rapidez y magnitud de las operaciones también entrañan desafíos. La IA acelera el bucle “observar-orientar-decidir-actuar”. Ese aumento de la rapidez y la magnitud puede dar lugar a percepciones erróneas debido a incoherencias entre las intenciones militares y los análisis generados por sistemas que funcionan con IA, lo que podría llevar a una intensificación involuntaria de las tensiones. La rapidez es también difícil de mantener el uso de la fuerza sujeto al juicio y el control humanos.

En este contexto, la Unión Europea destaca la importancia de la cooperación internacional en el estudio de la incidencia de la IA en el ámbito militar y de posibles marcos de gobernanza.

Anexo II

Replies received from international and regional organizations, the International Committee of the Red Cross, civil society, the scientific community and industry¹

A. International and regional organizations

African Commission on Human and Peoples' Rights

[11 April 2025]

I. Introduction

The African Commission on Human and Peoples' Rights (the African Commission), as the premier treaty-based human and peoples' rights body of the African Union (AU), is entrusted with the mandate of promoting and protecting human and peoples' rights in Africa under the African Charter on Human and Peoples' Rights (African Charter). In the African Commission's study on Addressing Human Rights Issues in Conflict Situations, the African Commission's Focal Point who led the study observed that 'it is ... in conflict and crisis situations that the most egregious violations and abuses of rights are perpetrated...With the changes in the nature of conflicts and the attendant heightened threat to human and peoples' rights, there is a greater need for the human rights system to pay increasing attention to and provide effective responses to the challenges that these new dynamics present to the protection and observance of rights.' In the current context, one of the major new dynamics that carries serious implications for peace and security and therefore human and peoples' rights relate to Artificial Intelligence (AI) and in particular its rapid development and use in the military domain.

During its 1214th meeting, the AU Peace and Security Council (PSC), in requesting the AU Commission to conduct a study to assess the adverse impact of AI on peace and security, underscored the necessity of ensuring African perspectives in shaping global AI governance frameworks. Against this background and having regard to its work on AI and other technologies and human and peoples' rights² and human rights in peace and security, the African Commission is pleased to share its views in response to the invitation of the Secretary-General for submission of inputs on AI in the military domain and its implications for international peace and security.³

II. AI in the military domain and peace and security

The development and use of AI technologies in the military domain particularly to automate military functions such as surveillance, targeting, and the deployment of

¹ In accordance with operative paragraph 8 of General Assembly resolution [79/239](#), the replies received from international and regional organizations, the International Committee of the Red Cross, civil society, the scientific community and industry are included in the original language received. The Secretary-General remains committed to multilingualism as a core value of the United Nations.

² Resolution ACHPR/Res. 473 (EXT.OS/ XXXI) 2021 on human and peoples' rights and artificial intelligence (AI), robotics and other new and emerging technologies in Africa, available at <https://achpr.au.int/en/adopted-resolutions/473-resolution-need-undertake-study-humanand-peoples-rights-and-art>.

³ The Focal Point of the African Commission on its study on human and peoples' rights and AI, robotics and other technologies acknowledges with appreciation the contribution of Professor Thompson Chengeta, who is the consultant providing technical assistance in the development of the study, through the Centre for Human Rights, University of Pretoria.

lethal force have far reaching consequences for peace and security and hence for human and peoples' rights. The AU Continental AI Strategy, endorsed during the 44th Extraordinary Session of the Executive Council of the African Union, highlights AI governance and regulatory challenges, particularly in military applications, warning that AI could exacerbate conflicts through inaccurate predictions or deployment of autonomous weapon systems. Additionally, the framework raises concern about disinformation, misinformation, cybersecurity threats, and military risks.

From the perspective of the development and use of AI in the military domain, peace and security should not be seen just from the perspective only of what it means for stability of states and societies. Beyond its conception under the UN Charter and public international law associated with friendly relations of states, peace and security is also a fundamental right of all peoples. The African Charter thus stipulates that 'All peoples shall have the right to national and international peace and security. The principles of solidarity and friendly relations implicitly affirmed by the Charter of the United Nations and reaffirmed by that of the Organization of African Unity shall govern relations between States.'⁴

The framing of peace and security as a right of peoples compels states to assess and govern the development and deployment of AI technologies in the military domain through a human rights lens that prioritises the prevention of harm, suffering, and injustice. Together international law conception of peace and security, it places an affirmative duty on states to ensure that AI systems do not contribute to conflict, perpetuate structural inequalities, or violate the rights and dignity of individuals and communities. By embedding peace and security within the framework of human rights, states are not only accountable for avoiding direct acts of aggression, but also for proactively creating and maintaining environments in which human flourishing, security, and justice are protected from the potentially disruptive or harmful impacts of emerging military technologies.

The implication of AI in the military domain to peace and security, farmed comprehensively, thus goes beyond how it shapes the obligation of states for non-aggression. It also covers how algorithm-driven systems may dehumanise individuals, introduce bias, and lead to unaccountable or disproportionate harm. It raises critical questions about the erosion of human oversight, the potential for unlawful killings or violations of international humanitarian law, and the targeting of vulnerable or marginalised populations.

By transforming military capabilities, the application of AI in the military domain can also have implications for peace and security by heightening tendencies for engaging in hostilities. The resultant escalation of tension and violence will be inimical not only to stability and peace between and within states but also most importantly carries more adverse consequences for the development needs of the less developed parts of the world such as Africa. While AI may contribute to advancing the development needs of Africa, its development and use in the military domain can have devastating consequences for development detrimental in particular to the right to development enshrined in Article 22 of the African Charter.⁵

This link between peace and development is also central to the Sustainable Development Goals (SDGs), especially SDG 16, which promotes peace, justice, and strong institutions. Without peace and security, sustainable development cannot be achieved. Recognising this link is critical in the governance of military AI, as the

⁴ Article 23(1) of the African Charter.

⁵ All peoples shall have the right to their economic, social and cultural development with due regard to their freedom and identity and in the equal enjoyment of the common heritage of mankind.

militarisation of AI can aggravate instability, particularly in fragile regions, and undermine Africa's developmental aspirations. By reaffirming the interconnectedness of peace and development, the African Commission calls for a governance approach that upholds peace as both a human right and a developmental imperative.

III. The need for a human and peoples' rights-based regulation of the development and use of AI in the military domain

Given the ways in which the use of AI in the military domain transforms the conduct of hostilities and how the development of AI relies on the extraction of natural resources particularly critical minerals such as rare earth minerals, it is the submission of the African Commission that both the process of extraction of resources in the development of AI in the military domain and the use of AI in the military domain need to be in full compliance with human and peoples' rights standards and international law principles, including international humanitarian law.

First and foremost, it is of paramount significance that the development and use of AI in the military domain complies with the right to peace and security enshrined in Article 23 of the African Charter on Human and Peoples' Rights. As a right that is born out of the recognition of the inseparability of the enjoyment of other human rights states from peace and security, this right entails that the use of AI in the military domain should be consistent with the international law prohibition of the use of force enshrined in the UN Charter and the Constitutive Act of the African Union.

Second, the use of AI technologies in conflict settings need to ensure respect for applicable human and peoples' rights and international humanitarian law principles, including most notably needs to adhere to the principles of precaution, necessity, distinction, proportionality and legitimacy. These requirements apply irrespective of whether the context in which the use of AI in the military domain relates to international armed conflicts or non-international armed conflicts. As established in the African Commission's study,⁶ parties to conflict are obliged to observe human rights standards where such conflicts do not meet the IHL threshold of armed conflict. As such, those who use AI technologies in conflict situations that do not meet the IHL threshold of armed conflict are legally obliged to respect and ensure respect for the human and peoples' rights standards established under treaty and customary international human rights law.

Third, the development of AI in the military domain and the use AI technologies in hostilities need to comply with the principle of transparency. This is fundamental because it is the basis for ensuring effective regulation of the development and use of AI in the military domain and for compliance with applicable human rights and international law standards. Additionally, transparency is critical for ensuring compliance with the obligation for respecting the dignity, privacy and data protection of individuals. The principle of transparency is also a pre-requisite for addressing some of the concerns that arise from use of AI in the military domain including bias (owing to the source and type of data used) and explainability. Transparency is also critical not only with the development of AI in the military domain but also with respect to the transfer of AI technologies in the military domain.

Fourth, from the perspective of human and peoples' rights and IHL, the other standard key to human rights and international law-based regulation of the development and use of AI concerns accountability. In the event of the occurrence of violations of human and peoples' rights standards or IHL principles from the development and use of AI in the military domain, there has to be both institutional and individual accountability. Accountability in this instance encompasses not only

⁶ ACHPR, Addressing human rights issues in conflict situations, <https://achpr.au.int/en/node/895>.

the measures that are taken against perpetrators but also the remedial steps that need to be put in place for redressing victims.

Fifth, building and sharing of technical knowhow critical to ensuring regulation by states is the other principle. Recent developments including the jamming of GPS systems affecting flights reported in Eastern DRC and the deployment by the Islamic State of West Africa of armed drones, highlight not only the need for effective regulation but also the need for developing the requisite infrastructure and technical capacity for ensuring effective regulation.

IV. The link between the development of AI in the military domain and Africa's natural resources and its implications for peace and security

The African Commission is also of the view that when discussing peace and security, stakeholders must be aware of the link between development of military AI, Africa's natural resources – particularly critical minerals – and the notion of peace and security. Article 21(1) of the African Charter on Human and Peoples' Rights affirms: "All peoples shall freely dispose of their wealth and natural resources. This right shall be exercised in the exclusive interest of the people. In no case shall a people be deprived of it."⁷ Article 21(5) further provides that "States parties to the present Charter shall undertake to eliminate all forms of foreign economic exploitation particularly that practised by international monopolies so as to enable their peoples to fully benefit from the advantages derived from their national resources."⁸

This provision is particularly important in the context of military AI, which depends heavily on critical minerals such as cobalt, lithium, and rare earth elements – resources abundantly found in Africa. The 2024 Report of the Chairperson of the African Commission's Working Group on Extractive Industries, Environment and Human Rights Violations, stressed the "significance of critical minerals for new and emerging technologies" and highlighted that Africa has been burdened by a "resource curse phenomenon."⁹ The report of the Chairperson noted that "extraction of minerals and other resources not only fuels but also at times becomes the site where contestation over whose control and use triggers conflicts. In some instances, this has created a vicious cycle of insecurity and violence, a condition that not only leads to major human and peoples' rights violations but also the perpetuation of a vacuum of effective governance and the concomitant exploitative, socially and environmentally costly extraction of the resources of the continent."¹⁰

Therefore, governance of military AI must not only ensure the legal use of force but also address the exploitative chains of extraction that power such technologies. This requires strict oversight, equitable benefit sharing, and regional solidarity to prevent Africa's resources from being used to fuel further conflict and inequality.

V. Conclusion

The African Commission is of the view that the development and use of AI in the military domain carries far reaching consequences for international peace and security in general and for less developed parts of the world such as in Africa that historically suffered violations and remain vulnerable to the adverse impacts of the development and use of AI in the military domain without robust and effective legal regime for such development and use in the military domain. The African

⁷ Article 21(1) of the African Charter.

⁸ Article 21(5) of the African Charter.

⁹ African Commission's Working Group on Extractive Industries, Environment and Human Rights Violations (2024), <https://achpr.au.int/en/intersession-activity-reports/extractive-industries-environment-and-human-rights-violations> (accessed 08 April 2025).

¹⁰ As above.

Commission affirms that the development and use of AI in the military domain needs to be regulated on the basis of international law, human and peoples' rights and international humanitarian law standards with particular regard to the development and peace and security interests and human and peoples' rights needs of less developed parts of the world.

More specifically, beyond and above the right to peace and security, the governance of AI in the military domain needs to ensure respect for applicable human and peoples' rights and international humanitarian law principles, including most notably needs to adhere to the principles of precaution, necessity, distinction, proportionality and legitimacy, the principles of transparency, accountability and redress for victims and the obligation to build and share technical knowhow necessary for enabling societies to avert the risks that the development and use of AI in the military domain carries for peace and security. Only by ensuring that the development and use of military AI are aligned with international legal standards including those relating to the right to peace and security, the right to development, the right to privacy and protection of personal data, the right to remedy and the responsibility for exercising human control, the right to and control over natural resources and by addressing the structural inequities underpinning global technological advancement, can states uphold their duties to their peoples and advance genuine peace, justice, and security in relation to the development and use of AI in the military domain.

B. International Committee of the Red Cross

[19 March 2024]

Summary

The full submission is available at: <https://www.icrc.org/en/article/artificial-intelligence-military-domain-icrc-submits-recommendations-un-secretary-general>.

The International Committee of the Red Cross (ICRC) welcomes the opportunity to submit its views for consideration by the United Nations Secretary-General, in accordance with resolution [79/239](#).

The recommendations that the ICRC makes in this submission are in line with its long-standing mandate and practice of promoting respect for and the development of IHL, including its application to new technologies of warfare. This submission is intended to support States in ensuring that military applications of AI comply with existing legal frameworks and, where necessary, identifying areas where additional legal, policy, or operational measures may be required.

1. Normative proposals: Reaffirming existing IHL as the starting point

The ICRC has consistently emphasized that, while IHL does not explicitly prohibit or regulate the use of AI in military applications, it does restrict its development and use, and places strict constraints on AI when it is integrated into weapon systems or used in some way to conduct warfare.¹

Existing and emerging normative proposals on the military application of AI should build upon established international legal frameworks and mechanisms, including IHL. Where necessary, these frameworks can be reinforced through the development of additional legal instruments, operational guidance or policy measures to address specific risks or challenges posed by emerging technologies. The form and content of such measures may vary depending on the specific use case. The ICRC

¹ This has also been affirmed by States, including in the UN General Assembly with Resolution [79/239](#).

encourages the international community to engage in concrete discussions on particular applications of AI in the military domain and to prioritize consideration of those that pose the greatest risks to people affected by armed conflicts.

2. A Human-centred Approach to military AI

In line with the resolution, the ICRC advocates for a human-centred approach to the development and use of AI in armed conflict.² This approach has at least two key dimensions: first, ensuring a focus on the humans who may be affected by the use of AI; and second, emphasizing the obligations and responsibilities of the humans using or ordering the use of AI in military operations.

Despite the growing development of AI-related technologies in the military domain, IHL requires individuals to make legal determinations. Humans must, for instance, determine the lawfulness of attacks that they plan, decide upon or execute, and they remain accountable for those determinations. The ICRC considers that human judgement is crucial for reducing humanitarian risks, addressing ethical concerns and ensuring compliance with IHL. Accordingly, while certain technical tasks may be carried out by machine processes, it is not the system itself that must comply with the law, but the humans using it.³

This does not mean that commanders and combatants cannot or should not use tools, including AI-decision-support systems. However, these tools must only be designed and used to support, rather than hinder or replace, human decision-making.⁴ Further, States and parties to armed conflicts must ensure that human control and judgement are preserved in decisions that pose risks to the life and dignity of people affected by armed conflict. This is essential for ensuring respect for applicable laws, including IHL, and upholding ethical standards.⁵

3. Specific Applications of AI in the military domain

The ICRC has identified three specific applications of AI in the military domain that pose particularly significant risks to those affected by armed conflict:

1. AI in Autonomous Weapon Systems

Resolution [79/239](#) acknowledges the increasing integration of AI into weapons and weapon systems, a development that raises significant legal and humanitarian concerns. The integration of AI, particularly machine learning (ML) techniques, into autonomous weapon systems (AWS) exacerbates existing challenges posed by AWS in ensuring compliance with IHL. In particular, it increases difficulties for human users to understand, predict, and control the system's functioning and effects.

Users of AWS must be able to, with a reasonable degree of certainty, predict the effects of that weapon in order to determine whether it can be directed at a specific military objective, and take steps to limit those predicted effects, as required by IHL. This entails the ability to understand the functioning of the AWS: the nature and functioning of its sensors, the definition of its target profile and the potential effects in the circumstances of use, including any risk of error or malfunction. This is

² ICRC, AI and machine learning in armed conflict: A human-centred approach, 2019 (updated in 2021).

³ ICRC, International Humanitarian Law and the Challenges of Contemporary Armed Conflicts: Building a Culture of Compliance for IHL to Protect Humanity in Today's and Future Conflicts (IHL Challenges Report), 2024, p. 61.

⁴ *Ibid.*; ICRC, IHL Challenges Report – Chapter 2: Contemporary and future challenges in the conduct of hostilities, 2019, p. 32.

⁵ ICRC, Decisions, Decisions, Decisions: computation and Artificial Intelligence in military decision-making, ICRC, 2024, p. 8.

particularly relevant for AWS that function in opaque ways (the “black box” challenge), such as AWS relying on AI techniques, which prevent the human user from being able to understand, predict or explain the system’s output. This impossibility effectively results in a lack of control over the weapon’s effects, rendering it indiscriminate by nature.

In this regard, we reiterate the joint call made by the ICRC President, with the UN secretary-general,⁶ for new, legally binding rules prohibiting certain AWS and constraining the use of others.⁷ In particular, we recommend a prohibition on

- unpredictable autonomous weapons – those that, due to their design or the circumstances and manner of use, do not allow a human user to understand, explain or predict the system’s functioning and effects;
- autonomous weapons designed or used to target humans directly. This is required because of the significant risk of IHL violations and the unacceptability of anti-personnel autonomous weapons from an ethical perspective.⁸

The ICRC supports all efforts by States to urgently adopt a legally binding instrument to regulate AWS, in whichever forum they choose.⁹ The integration of AI into AWS should also be considered when discussing normative proposals on military applications of AI. Doing so is essential to ensure a consistent and comprehensive approach to the regulation of military AI, to avoid normative gaps, and to effectively address the serious legal, ethical, and humanitarian risks that are exacerbated by the integration of AI into AWS. In this regard, the ICRC considers it important that binding prohibitions and restrictions on AWS, including AWS that incorporate AI, are integrated into broader discussions on the governance of military AI.

2. *AI in Military Decision-Making*

AI decision-support systems (AI-DSS) are computerised tools that bring together data sources – such as satellite imagery, sensor data, social media feeds or mobile phone signals – and draw on them to present analyses, recommendations and predictions to decision makers.

The use of AI-DSS raise concerns related to system functioning, data quality, and human-machine interaction. These systems risk increasing the rate of unforeseen errors, perpetuating problematic biases – particularly those based on age, gender, ethnicity, or disability, and making it difficult for the users to understand how and why the system generates its output from a given input.

Generally, AI-based systems will perform better when given well-defined goals and access to representative and high-quality data. However, armed conflict environments are marked by uncertainty, volatility, and deliberate deception techniques by adversaries, which makes it extremely difficult to obtain reliable or transferable data. Even where good data exists, it may not reflect the specific operational or humanitarian dynamics of a particular context.¹⁰ Moreover, for AI systems that rely on training data, the utility of those data can rapidly diminish once

⁶ ICRC, Joint call by the United Nations Secretary-General and the President of the International Committee of the Red Cross for States to establish new prohibitions and restrictions on Autonomous Weapon Systems, 2023.

⁷ ICRC, ICRC Submission on AWS to the UN Secretary-General, 2024, p. 6.

⁸ *Ibid.*

⁹ *Ibid.*

¹⁰ ICRC, IHL Challenges Report, 2024, pp. 64–65; ICRC, AI and machine learning in armed conflict: A human-centred approach, 2019 (updated in 2021); ICRC, Decisions, Decisions, Decisions: Computation and Artificial Intelligence in Military Decision-Making, ICRC, 2024, pp. 31 and 54.

a conflict begins. Parties to armed conflicts will continuously seek to maintain the initiative and operate in a manner that is not anticipated by their adversary, adapting their strategies and tactics accordingly. This can fundamentally alter the environment in which the system was expected to operate, making the original data no longer representative of the new operational conditions. In such cases, the system's outputs may become unreliable, and the AI model may require re-evaluation or retraining in order to remain fit for purpose.

Human interaction with these systems raises further concerns, such as “automation bias” – a propensity to rely on machine outputs even when other available information may call those outputs into question – which is particularly pronounced in high-pressure or stressful environments like in armed conflicts.¹¹ Taken together, these factors can hamper a user’s ability to scrutinize the information available. The practical consequence might be, for instance, that someone plans, decides upon or launches an attack based solely on an AI-DSS’s output, thereby effectively serving as a human rubber stamp rather than assessing the lawfulness of the attack by considering all the information reasonably available including the AI-DSS output.¹²

On the positive side, the careful use of AI-based systems may facilitate quicker and more comprehensive information analysis, which can support decisions in a way that enhances IHL compliance and minimizes risks for civilians. In the context of urban warfare in particular, the ICRC has recommended that online open-source repositories should be used to gather information about the presence of civilians and civilian objects.¹³ Importantly, IHL imposes obligations to take constant care to spare the civilian population and to take all feasible precautions in attack. Therefore, in developing and using AI-DSS, armed forces should be considering not only how such tools can assist them to achieve military objectives with less civilian harm, but also how they might be designed and used specifically to protect civilians. However, the important point is that these computer outputs can inform but must not displace the need for legal determinations.

Beyond targeting decisions, militaries are also exploring the use of AI to support other operations traditionally carried out by humans, including detention operations. While technology deployed responsibly and with robust human oversight can contribute to IHL compliance, it also carries risks including bias, lack of transparency, and faulty programming and analysis, all of which can undermine compliance with IHL.¹⁴

To support efforts by States and other actors to ensure that military uses of AI-DSS remain consistent with IHL and humanitarian principles, the ICRC has formulated a non-exhaustive set of preliminary recommendations relating to the development and use of AI-DSS in armed conflict. They focus on 1) ensuring human control and judgement; 2) system design requirements; 3) testing, evaluation, verification and validation; 4) legal reviews; 5) operational constraints on use; 6) user training; 7) after-action reviews; and 8) accountability, among others. The recommendations are annexed to the full version of this submission.

¹¹ ICRC and the Geneva Academy, Artificial Intelligence and Related Technologies in Military Decision-Making on the Use of Force in Armed Conflicts: Current Developments and Potential Implications, ICRC, 2024, p. 17.

¹² ICRC, IHL Challenges Report, 2024, p. 65.

¹³ *Ibid.*, p. 66; ICRC, Reducing Civilian Harm in Urban Warfare: A Handbook for Armed Groups, 2023, p. 15.

¹⁴ ICRC, IHL Challenges Report, 2024, p. 22.

3. AI in Information and Communications Technologies

AI is expected to change how actors defend against and conduct information and communications technology (ICT) activities, including in armed conflict. In particular, States have noted with concern that the use of AI and other emerging technologies in malicious ICT activities may further increase their scale and speed, as well as the harm they may cause.¹⁵ For example, AI enables tools to identify and develop exploits for new vulnerabilities in software or networks, or to conduct harmful ICT activities autonomously, whether in offence or in defence. The ICRC is concerned that this could increase the risks of indiscriminate attacks, incidental civilian harm, including damage to critical civilian infrastructure, as well as the uncontrolled escalation of conflict, particularly in complex and interconnected digital environments.¹⁶

Similarly, information or psychological operations are not a new feature of armed conflicts; however, AI is changing how information is created and spread. AI-enabled systems, particularly generative AI, have been widely used to produce harmful content – text, audio, photos and video – which is increasingly difficult to distinguish from authentic, original content.¹⁷ The ICRC is concerned about the consequences for civilians that might result from the creation and spread of such information through ICT, including information that contributes to or encourages violence, causes lasting psychological harm, undermines access to essential services or disrupts the operations of humanitarian organizations.

In light of these concerns, the ICRC underlines the importance of applying existing international law, including IHL, to the use of AI in ICT activities. The ICRC urges States to ensure that the development and use of AI-supported ICT activities respect the protections afforded to civilians and civilian infrastructure in armed conflict. Moreover, in light of the emergence of increasingly autonomous ICT capabilities, the ICRC further encourages States to address the serious challenges posed by these tools, particularly by considering whether existing international law, including IHL, provides sufficient safeguards against the harm such tools can cause, or whether additional limits are needed.

4. Conclusion

The ICRC is grateful for the opportunity to share its above views and recommendations on ways to address the challenges and concerns raised by AI for the secretary-general's consideration, and stands ready to contribute further to assist States in taking effective action to address the risks posed by AI applications in the military domain.

C. Civil society

Autonorms

[10 April 2025]

The following is the AutoNorms project's submission pursuant to Resolution [79/239](#) on “Artificial intelligence in the military domain and its implications for international peace and security” adopted by the United Nations General Assembly on 24 December 2024. The resolution requests the UN Secretary-General to seek

¹⁵ 34th International Conference of the Red Cross and Red Crescent, Resolution 2 “Protecting civilians and other protected persons and objects against the potential human cost of ICT activities during armed conflict”, 2024.

¹⁶ ICRC, IHL Challenges Report, 2024, pp. 66–67.

¹⁷ *Ibid.*, pp. 58–59.

views, including those of Member States, civil society, the scientific community and industry, on “opportunities and challenges posed by the application of artificial intelligence in the military domain, **with specific focus on areas other than lethal autonomous weapons systems**”. The AutoNorms team welcomes the opportunity for representatives of academia to submit their views on this important and timely topic.

The AutoNorms project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement no. 852123). Led by Professor Ingvild Bode and hosted by the Center for War Studies at the University of Southern Denmark, the project examines how the integration of artificial intelligence (AI) technologies into weapon systems and military targeting shapes international norms governing the use of force.¹

Introduction

Over the past 2-3 years, the international debate about applications of AI in the military domain has been characterized by two significant, near-simultaneous changes. First, there has been a **move away from its predominant focus on autonomous or AI technologies in weapon systems** towards considering AI technologies across a wider range of military decision-making tasks, especially in relation to targeting. To reflect his move, this submission focuses on the employment of **AI-based decision support systems (AI DSS)**, or systems that are meant to be used as tools to directly or indirectly inform the complex process of use-of-force decision-making, for example, by analyzing large volumes of data, recognizing patterns within the data, predicting scenarios, or recommending potential courses of action to human decision makers.

Second, there has been a **growing emphasis on human-machine interaction** in the context of using AI in the military domain.² This emphasis results from the broad recognition that, even when humans are ‘in’ or ‘on’ the loop of targeting decision-making, they need to exercise a sufficient level of oversight, control, and agency over the targeting process. Human oversight is a governance principle featuring prominently across various international initiatives, including [A/RES/79/239](#). However, **dynamics of human-machine interaction as part of the use of AI DSS both introduce new issues and solidify existing sets of challenges that require governance attention**. Our submission highlights these challenges and the need to ensure the exercise of human oversight and agency throughout the full targeting decision-making spectrum. It is structured in three parts, starting with explicating challenges of human-machine interaction, then commenting on the relative under-development of the international debate about AI DSS, and finally, sketching a way forward.

Challenges of human-machine interaction in the use of AI DSS

The use of AI DSS involves various dynamics of human-machine interaction because military personnel such as operators and intelligence analysts routinely and increasingly interact with a network of AI systems throughout the targeting process. These interactions involve multiple challenges **which have the potential to affect**

¹ The members of the AutoNorms team are Professor Ingvild Bode, Dr Hendrik Huelss, Dr Anna Nadibaidze, Dr Guangyu Qiao-Franco, and Dr Qiaochu Zhang. The AutoNorms project is based at the Center for War Studies, University of Southern Denmark, Odense, Denmark. For more information, please visit our website: www.autonorms.eu.

² Ingvild Bode and Anna Nadibaidze, “Symposium on Military AI and the Law of Armed Conflict: Human-Machine Interaction in the Military Domain and the Responsible AI Framework,” *Opinio Juris*, April 4, 2024, <https://opiniojuris.org/2024/04/04/symposium-on-military-ai-and-the-law-of-armed-conflict-human-machine-interaction-in-the-military-domain-and-the-responsible-ai-framework/>.

the exercise of human agency, or humans' capacity to understand a system's functions and its effects in a relevant context; deliberate and decide upon suitable actions in a timely manner; and act in a way where responsibility is guaranteed.³

Dynamics of human-machine interaction result in **distributed agency between humans and AI systems, where they are not separated into two distinct entities but rather form part of a socio-technical system**.⁴ As part of this system, both sides may influence each other in different ways, which then translate into various forms of distributed agency located along a spectrum. In some instances, dynamics of human-machine interaction will offer more opportunities for exercising human agency in targeting decisions. In other instances, however, the humans involved in use-of-force decision-making will be more constrained in their ability to exercise agency.

For example, **humans' ability to exercise agency might be limited by cognitive biases such as automation bias or anchoring bias**. Humans could over-trust AI DSS even when knowing that there might be malfunctions or unintended errors involved, risking an overreliance on algorithmic outputs without engaging in the critical deliberations and assessments that are needed to exercise human agency, especially in critical targeting decisions that might inflict death, destruction, and severe harm. Such biases are typically exacerbated by the increased speed of AI-assisted military decision-making, especially in contexts where there are high levels of pressure to act rapidly. They can also be exacerbated by AI DSS that are used for prescription or recommendations, because such systems restrict the options or courses of action available to human decision makers.

Moreover, given that AI DSS are likely to be employed not individually but rather as part of a network of systems, the increased complexity of interactions can result in situations where humans act upon some outputs suggested by AI DSS, but do not overall exercise a high quality of agency. Due to these and many other concerns related to interactions between humans and AI DSS, **there is a need to further investigate challenges of human-machine interaction that result in AI DSS not positively ‘supporting’ humans but rather undermining humans’ ability to exercise agency**.⁵

The risks of not addressing challenges of distributed agency are substantial. First, situations where humans are restricted in their exercise of agency **raise questions about compliance with international humanitarian law**, which requires that humans be held accountable and legally responsible for violations of legal principles. Although humans remain officially in control of the selection and engagement of targets, there are concerns about the exact role played by humans in context of using AI DSS in practice.

Second, these concerns also extend to the risk of **negatively affecting moral agency and responsibility in warfare**. Challenges of human-machine interaction that result in distributed agency would allow humans to feel less morally responsible

³ Anna Nadibaidze, Ingvild Bode, and Qiaochu Zhang, *AI in Military Decision Support Systems: A Review of Developments and Debates* (Odense: Center for War Studies, 2024), <https://www.autonorms.eu/ai-in-military-decision-support-systems-a-review-of-developments-and-debates/>.

⁴ Ingvild Bode, *Human-Machine Interaction and Human Agency in the Military Domain*, Policy Brief No. 193 (Waterloo, ON: Centre for International Governance Innovation, 2025), <https://www.cigionline.org/publications/human-machine-interaction-and-human-agency-in-the-military-domain/>.

⁵ Anna Nadibaidze, “Do AI Decision Support Systems ‘Support’ Humans in Military Decision-Making on the Use of Force?” *Opinio Juris*, November 29, 2024, <https://opiniojuris.org/2024/11/29/do-ai-decision-support-systems-support-humans-in-military-decision-making-on-the-use-of-force/>.

for decisions that could affect other people's lives. They also risk making the human role a nominal, 'box-checking' exercise which can *de facto* be compared with AI DSS playing an 'autonomous' role because the human role is substantially reduced.

Third, there are **security and operational risks related to distributed agency dynamics**, especially when they give too prominent roles to AI DSS and algorithmic outputs. AI systems often malfunction, are trained on biased sets of data which do not apply beyond the training context or specific contexts of use, as well as integrate assumptions that might not be strategically or operationally beneficial.

Various types of biases, issues of trust, uncertainties, targeting and military doctrines, political and societal contexts in which AI DSS are used – all these aspects **can lead to dynamics of distributed agency which limit the exercise of human agency and prioritize algorithmic outputs**. It is important to investigate these dynamics and ensure that distributed agency provides more opportunities than limitations to human decision makers in warfare.

Relative under-development of the international debate on AI DSS

Despite increasing reports about the use of AI DSS in recent and ongoing armed conflicts, and the significant challenges and risks they pose to the effective exercise of human agency, **the international debate on human-machine interaction in the use of AI DSS remains insufficiently developed**, particularly within intergovernmental UN settings. Current discussions on AI in the military domain, including those within the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (GGE on LAWS), have focused on the use of AI at the tail-end of the targeting process, specifically autonomy and AI in weapon systems. This narrow focus risks overlooking or failing to address critical normative, legal, ethical, security, and operational risks that can proliferate and compound throughout the entire targeting decision-making process.

An increasing, albeit still limited, number of stakeholders are raising this issue at international multistakeholder forums, such as the Summits on Responsible Artificial Intelligence in the Military Domain (REAIM). Some international non-governmental organisations and research institutes – such as the International Committee of the Red Cross (ICRC), the Stockholm International Peace Research Institute (SIPRI), the UN Institute for Disarmament Affairs (UNIDIR), and the Asser Institute – have initiated discussions on challenges posed by AI in the military domain, beyond the issue of autonomy in weapon systems. Despite this progress, there remains **a clear need to develop a more comprehensive and inclusive international multistakeholder debate to guide the responsible development and deployment of AI DSS in military contexts**.

Way forward

In closing, we sketch three ways intended to move the international debate about applications of AI in the military domain forward:

1. **Increase awareness for the implications of practices of designing, developing, and using AI DSS.** States and other stakeholders across industry, civil society, and academia engaged in the governance, development, and use of AI DSS for military targeting must consider the implications of their practices. These practices influence what counts as 'appropriate' ways of considering and employing AI DSS and thereby shape what becomes the accepted, requisite quality of human oversight and agency exercised over the whole process of use-of-force decision-making. To increase such awareness, the debate pursuant to [A/RES/79/239](#) at the UNGA First Committee should centrally focus on the issue of AI DSS in the military domain.

2. Consistently map both ‘best’ practices and ‘problematic’ practices associated with the design, development, and use of AI DSS. To get a better sense of the direction that the design, development, and use of AI DSS take, states and other stakeholders need to closely map their own (and others’) practices. While there have been some limited efforts to exchange potential best practices, we also need to be attentive to practices with potentially problematic effects. This should encompass practices exercised across the full life cycle of AI systems from development to use and post-use review. Mapping such practices would offer stakeholders a better overview of which practices may be beneficial, i.e., provide opportunities for the exercise of human agency, and which practices may be problematic, i.e., limit the exercise of human agency, and therefore assess the desirability of particular practices.

3. Pursue the debate on AI DSS within a multistakeholder format. States should work with diverse stakeholders – including academics across social sciences and technical disciplines, civil society representatives, and international organizations – to develop normative guidance and regulation, especially regarding the human role in military decision-making. Moreover, top-down processes towards governing AI DSS should be accompanied by a bottom-up, standard-setting process focused on establishing operational standards. Such an inclusive approach could strike a balance between national security and humanitarian concerns, while reinforcing the need to ensure that humans can exercise agency in use-of-force decisions.

Global Commission on Responsible Artificial Intelligence (AI) in the Military Domain

[11 April 2025]

1. Introduction

The Global Commission on Responsible Artificial Intelligence (AI) in the Military Domain (GC REAIM) welcomes the opportunity to contribute to the United Nations Secretary-General’s report pursuant to resolution [A/RES/79/239](#).

GC REAIM recognises that military applications of AI present both opportunities and challenges for global peace and security. Accordingly, the establishment of responsible and ethical governance – consistent with States’ obligations under applicable international law – is essential. The global community must take proactive steps to ensure that military AI is developed and deployed in a manner that de-escalates rather than escalates conflicts; respects and enhances, rather than compromises, the sovereignty and territorial integrity of states; promotes rather than threatens the security and safety of civilians; constrains and supports rather than erodes the existing rules-based international order.

In line with GC REAIM’s resolute commitment to advancing international governance efforts, this note outlines some of the – non-exhaustive – views expressed by GC REAIM Commissioners and Experts on the implications of AI in the military domain to peace and security. The views presented are general in nature and will be further elaborated in the forthcoming GC REAIM report. While the Commission plans to present substantive and actionable recommendations for stakeholders in September 2025, this note does not yet include concrete proposals. As discussions among Commissioners and Experts are still ongoing, it instead highlights some of the key opportunities, challenges, benefits and risks posed by AI in the military domain to peace and security.

2. Technological Foundations

GC REAIM holds that meaningful policy deliberations on AI in the military domain must be grounded in a shared, foundational understanding of the underlying technologies and their potential trajectories. The complexity of AI technologies often gives rise to misunderstandings, inflated expectations, or misguided applications. Consequently, it is imperative to demystify AI through formal and well-defined frameworks that distinguish between current capabilities and speculative future developments. To support this objective, GC REAIM is developing a taxonomy which seeks to map the full spectrum of AI applications across military and broader peace and security contexts. The taxonomy differentiates between the implications of AI in operational activities – such as warfighting and intelligence – and administrative activities – such as logistics and personnel training and helps identify the specific applications of AI that should be prioritised in governance deliberations.

In its approach to the creation of a taxonomy, GC REAIM highlights the need for and contributes to a concerted effort to clarify, standardise, and encourage the accurate use of technical language with different layers of abstraction for policymakers, experts, and the public, thereby enhancing transparency, mutual understanding, and public trust. GC REAIM also cautions against the uncritical multiplication or adoption of new terminologies in AI governance discourse, unless these are clearly defined; and to ensure such terms are not used to circumvent or obscure existing legal obligations. Precision and consistency in language are the basis of responsible AI governance.

3. Implications for Peace, Security, and Stability

GC REAIM recognises that the integration of AI into the military domain presents benefits as well as both foreseeable and unforeseeable risks to international peace and security. A balanced approach to the range of opportunities and challenges emerging throughout the AI life cycle lies at the core of GC REAIM's method and is essential for responsible AI governance.

AI in the military domain may contribute to international peace and security in several important ways. At the developmental stage, the advancement of military AI capabilities may act as a deterrent to violence, as the mere development and presence of advanced technologies by responsible actors can encourage restraint by aggressors. Military AI may enhance early warning systems, strengthening conflict prevention strategies, and supporting arms control verification through AI-driven tools that foster transparency, trust, and cooperation among states – fundamental elements in conflict prevention. AI can also bolster national security and defence by improving the precision, accuracy, and efficiency of intelligence analysis and situational awareness, enabling real-time threat detection, and facilitating more efficient counterterrorism operations through predictive analytics and autonomous systems. AI-powered systems can rapidly process vast amounts of complex data, enabling military forces to make timely, informed decisions that may prevent escalation and support conflict de-escalation efforts. These traits can also help improve targeting accuracy and precision, potentially reducing the risk of collateral damage or fratricide – attacks on one's own forces – and aiding compliance with International Humanitarian Law (IHL) to protect the security of protected persons, such as civilians and non-combatants, during armed conflict. Military AI may also reduce certain forms of human bias and enhance accountability by providing precise data, surveillance, and real-time monitoring, enabling clear attribution of actions to specific actors. In these ways, AI offers meaningful opportunities to reinforce adherence to international law and ethical standards, strengthening the normative foundation of the rules-based international order underpinning global peace and security.

AI in the military domain also presents a range of risks. As with the development of other general-purpose technologies, the development of AI in the military domain may accelerate arms races. AI technologies driven by the commercial market may be repurposed by militaries or soldiers in need or increase the access of violent non-state actors to AI-enabled military capabilities, which may intensify ongoing conflicts and contribute to broader instability. There are also concerns that states could employ AI technologies to suppress human rights, entrench internal repression, and destabilise both regional and global peace.

Concurrently, as with AI more broadly, the environmental consequences of military AI – such as the energy-intensive demands of AI systems, resource extraction, and ecological damage from AI-enabled military systems – could aggravate resource scarcity and environmental degradation, fuelling tensions and undermining long-term peace. However, given the impact militaries have on civilian technology development, efforts to reduce the environmental impact of AI in defence settings could have far-reaching beneficial consequences for all uses of AI. As such, considerations of environmental impacts should be a component of responsible AI governance in the military domain.

The large-scale data extraction required for AI development could intensify geopolitical rivalries, facilitate intrusive surveillance, and create distrust through opaque and exploitative data practices. Such deployment of military AI may perpetuate discrimination and exacerbate social divisions, undermining stability and ultimately international peace and security.

There are simultaneously significant concerns regarding the potential of integration of AI within the command, control, and communication (C3) structures of nuclear weapons. A number of Commissioners and Experts have emphasised that this is a red line that must not be crossed. The commitment of several nuclear-armed states to human decision-making surrounding the employment of nuclear weapons is therefore applauded. Further, the development of large-scale lethal autonomous weapon systems – such as swarms of anti-personnel devices – risks creating a new category of weapons of mass destruction, posing serious threats to global peace and security. Relatedly, AI may lower the barriers to creation and use of nuclear, chemical, or biological weapons by state or non-state actors, thus generating new challenges for arms control and non-proliferation regimes.

Beyond these strategic risks, AI may affect the character of war and lower the thresholds for armed conflict. By increasing the speed of armed escalation and driving changes in the capabilities of weapons systems, AI in the military domain may reduce states' confidence in their deterrent capabilities – particularly in the face of cyber infiltration risk – thus influencing how decision makers receive, process, and act on information. AI in the military domain could also exacerbate asymmetric warfare and violence by widening technological disparities that could increase the likelihood of force being used prematurely or disproportionately.

Operationally, inaccurate AI systems used for targeting can undermine the security of protected persons under IHL by increasing the risk of indiscriminate attacks, violations of proportionality, and failure to distinguish between combatants and civilians. Closely related to this is the risk of fratricide due to potential errors in target identification or decision-making, which can undermine operational effectiveness, escalate conflict, and erode trust within militaries and alliances. Finally, there are views that the use of certain AI systems in the military domain can create accountability gaps absent clear rules. By complicating the attribution of responsibility for unlawful actions, the deployment of AI in the military domain could undermine key principles of international law and state responsibility for internationally wrongful acts. This may complicate efforts to hold individuals or

states responsible for violations, leading to a reduced deterrent effect against unlawful conduct. Without avenues to hold actors legally responsible, the enforcement of international law weakens, potentially destabilising peace, encouraging impunity, and exacerbating global insecurity.

4. Decision-Making and Responsibility

GC REAIM acknowledges the ethical and legal challenges that arise from integrating AI into military decision-making which may have a direct impact on preservation of peace and security. The relationship between human judgment and machine outputs is complex and without measures to ensure lawful, responsible and effective development and deployment, there can be an erosion of accountability and increased risks of unintended harm. As AI systems become more sophisticated and integrated within military capabilities, it is plausible that algorithmic decisions may become more commonplace across global battlefields, introducing moral and legal challenges regarding human control, oversight and judgment in diverse contexts.

To address these risks, GC REAIM promotes the need for context-appropriate human judgement over specific uses, capabilities and decisions of AI in military applications. The GC REAIM report will list considerations and conditions that underpin and support human responsibility, judgment and means of adequately evaluating relevant actions and decisions. This could include the introduction of technical standards for explainability, as well as maintaining appropriate human oversight in targeting decisions, assessments of precautions, proportionality and distinction, and other critical operational choices. However, given that the very definition of autonomy in machines suggests the minimisation or removal of the human, ensuring human responsibility and accountability may require focusing on human decision-making at earlier stages of a system's life cycle, as the systems structure the behaviour of all who work with it. Human oversight is essential to uphold state obligations under applicable international law, in particular, IHL.

Military AI systems must be designed not only to support all individual and collective agents in the military domain to be effective in safely carrying out their lawful tasks, but also to do so responsibly and without compromising or undermining their status as moral human agents. GC REAIM suggests that military AI based socio-technical systems need to be explicitly and demonstrably designed to adequately attribute and apportion responsibilities and is determined to contribute to this process. For the security of protected persons, parties to armed conflicts should at all times be able to demonstrate that everything possible has been undertaken to create the conditions under which military personnel can effectively apply extant and widely shared principles and laws of armed conflict to their own situation, when using or relying upon AI components in the execution of their tasks.

5. Governance and Regulation

In light of both the opportunities and risks associated with military AI, GC REAIM supports a comprehensive governance framework that implements authentic international law. GC REAIM reiterates that existing legal regimes provide a solid foundation for regulating AI technologies. Governance must incorporate and account for procedural safeguards (due diligence and legal reviews, transparency of testing, evaluation, and validation, accreditation, and verification), substantive obligations drawn from various branches of international law, and soft law tools (military doctrines, national policies and strategies, norms and standards). In principle, all relevant international legal frameworks must be considered and applied. These include, but are not limited to, the following: (1) international law (*jus ad bellum*) which regulates when and how states use force, codifying a general prohibition on the use of force and exceptions such as in the case of self-defence, (2) international

humanitarian law (*jus in bello*) which governs conduct during armed conflict and ensures the security of protected persons, (3) international human rights law.

GC REAIM further emphasises the critical role of international, regional, and domestic institutions in implementing and enforcing these legal norms. Effective governance requires collaboration across these levels and the inclusion of both binding (hard law) and non-binding (soft law) instruments. Soft law mechanisms, such as codes of conduct and ethical principles, can complement existing treaties and facilitate rapid, flexible responses to technological developments.

To address the diverse range of challenges surrounding the integration of AI into the military domain, GC REAIM supports proactive risk-mitigation and confidence-building measures. While binding regimes are challenging for general purpose technologies, there may be opportunities for rigorous monitoring, verification, and enforcement mechanisms inspired by successful global arms control regimes. For example, Commissioners and Experts have discussed ideas such as an Autonomous Incidents Agreement to reduce the risks of miscalculation among AI-enabled autonomous systems, or a committee or consortium that could set guidelines and recommendations surrounding the testing and evaluation of AI systems, including generative AI. GC REAIM also suggests that states and industries should consider adopting human-centred safety-by-design principles, implement red-teaming practices throughout AI system life cycles, and maintain clear chains of accountability for all actors. Only through robust multilateral dialogue and inclusive multi-stakeholder cooperation can AI be effectively governed to enhance peace and security rather than exacerbate global instability.

GC REAIM acknowledges that the development of a comprehensive governance framework for military AI faces several key challenges. First, there is the challenge of diverse interests and perspectives, with states, private companies, and civil society holding varying and sometimes conflicting views on the regulation of military AI. Second, the sensitivity surrounding national security and defence poses a significant barrier, as many states are reluctant to subject their military technologies to international scrutiny or regulation due to legitimate security interests. Third, achieving meaningful and substantive inclusivity in discussions is often difficult, as key stakeholders may be excluded or marginalised in decision-making processes. Fourth, a trust deficit between states, international organisations, and the private sector complicates efforts to establish cooperative governance. Fifth, the presence of crosstalk, incommensurability, and discursive dissonance arises due to the diverse backgrounds and expertise of stakeholders, making consensus-building challenging. Finally, these obstacles are compounded by the lack of clear frameworks that address the complex ethical, legal, and technical issues at the nexus of AI and the military domain. In light of these challenges, the final GC REAIM report will offer strategies to navigate and overcome these barriers in developing a robust governance framework.

6. Conclusion

GC REAIM observes that the rapid advancement and deployment of AI technologies in military contexts poses opportunities, challenges, benefits and risks for global peace and security. Balancing these considerations must be met with a technologically sound, inclusive, principled, and legally grounded approach to governance.

A clear understanding of AI's technological foundations is necessary to properly address its role in modern warfare. Ethical and legal responsibility should remain human-centred, and governance frameworks must rely on the robust application of international law, supplemented by cooperative multilateral efforts and soft law

instruments when appropriate. In its formation and deliberations, GC REAIM has had the opportunity to reflect upon the conversations happening in broader governance processes, finding ways to effectively bridge gaps between disciplines and regional perspectives.

GC REAIM urges the United Nations and all State Parties to place these principles at the heart of global discussions on the implications of AI in the military and broader peace and security, for the present and future generations. Only through concerted international cooperation, guided by a shared commitment to human dignity, peace, and justice, can we ensure that the future of AI in the military domain is one that strengthens our common security.

InterAgency Institute

[11 April 2025]

The InterAgency Institute was established in December 2020 as a digital think tank, founded by expatriate and Global South women as a collective of researchers. It is in this condition that we address this submission on “opportunities and challenges posed to international peace and security by the application of artificial intelligence in the military domain, with specific focus on areas other than lethal autonomous weapons systems,” following [A/RES/79/239](#). With this, we seek to craft a complementary set of suggestions to develop the policy discussion in points where understanding that AI encompasses a wide array of data-processing techniques, and may be integrated into different types of warfare, in multiple parts of the organization, and at different levels.

The InterAgency Institute would like to point to overarching trends that fall within our areas of expertise, namely: (1) a focus on the global south, specially in how to prevent furthering the security gap; and (2) in how interagency cooperation in a time of greater mistrust may be leveraged to ensure the integration of AI in the military does not. Additionally, we make the point that Decision Support Systems (DSS) create analogous problems when compared to Autonomous Weapon Systems (AWS).

1. Addressing the security gap between the Global South and the Global North

The increasing technological intensity and digitalization of the battlefield are likely to increase the capacity gap between countries in the Global North & South. The “optimization of war” entails furthering this discrepancy, augmenting threats, and deteriorating the global security landscape. The wide range of AI-enabled solutions represents discrepant utility levels across tools.

While some tools require a low threshold (thus providing usually an equally low ceiling), the systems that pose the biggest military advantage require a high knowledge threshold to be implemented, therefore, will likely not be open source, and will only be available to entities with sufficient means to develop or acquire them. Given the experience in past decades on multilateral forums, it is important to recognize that interest in access to these technologies will play a role in the negotiations.

In the long term, the current trend of “technological sovereignty” (or more specifically of restricted technological access due to global inequalities) may be transformed to undermine such technological control, creating far-reaching implications of this new revolution in warfare, involving stakeholders that may be reluctant to shape modern discussions due to a lack of current development of these technologies in their ecosystem.

2. InterAgency cooperation in times of distrust

These issues call for interagency cooperation at both the strategic and operational levels. The lack of interagency cooperation might lead to threat escalation and the eroding foundations for peace and security. Interagency cooperation should focus on formalizing specific channels for communication between different States, developing strategies for AI implementation that will not damage diplomatic relations, and generating more transparency in the interactions between agencies and contractors. The participation of different branches of government at the UN-level discussions is pivotal for a whole-of-government perspective in the deliberations. Beyond interagency cooperation at the governmental level, the wide array of applications of military AI calls for different sets of Confidence Building Measures (CBMs).

Since AI may be integrated in different warfare types and at different levels, its applications for different contexts have different ethical implications and consequences. Therefore, a monolithic understanding of risks posed by AI in the military context and consequently a unique set of CBMs would be inadvisable. CBMs for AI use in the strategic level of cyberspace will not be the same as CBMs for AI use in the tactical level of aerial warfare. Therefore, thinking about CBMs for military AI as a monolith will lead to inaccurate and in some cases inapplicable measures, undermining its effectiveness.

There is a necessity for sharing best practices in the introduction of AI into military procedures. In this sense, a trade-off should be made, prioritizing best practices that contribute to strengthening the aforementioned points of interagency cooperation and CBMs, and other practices that fall within the larger umbrella of strengthening international peace and security. Sharing of best practices relating to cybersecurity and reliability of the technology could also take place, but they should give priority to CBMs that focus on integration of AI at the strategic level and in manners that avoid the escalation of threats.

3. Decision Support Systems

Target identification or recognition via AI-enabled Decision Support Systems (DSS) entail analogue problems to Autonomous Weapon Systems (AWS). Digital dehumanization, lowering the threshold of violence, and automation bias are byproducts of that process that may only be avoided by the creation of red lines prohibiting such systems that replicate those concerns.

This problem stems not only from AI, but from a wider trend. Other data processing techniques that involve deterministic sorting of data that is not adequately processed by human operators also generate these problems. This caveat should be made to understand that not only systems with AI-enabled technology in DSS pose these kinds of threats, but a wider array of data gathering/processing techniques.

Conclusions and recommendations

- Formal interagency bodies to interface with multilateral AI/military tech negotiations
- Funding and support for academic research in the Global South focused on military AI implications;
- Regular technical-diplomatic summits focused on transparency, shared definitions, and threat perception;
- Prioritize capacity-building initiatives for Global South actors;

- Red lines and confidence building measures could be tailored to the specific technology and operational context;
- The discussions on Autonomous Weapon Systems encapsulate worries around AI-enabled Decision Support Systems. The creation of red-lines for these systems could benefit from building upon recommendations of the GGE on LAWS;

International Committee for Robot Arms Control

[11 April 2025]

The International Committee for Robot Arms Control (ICRAC) values the opportunity to submit our views to the United Nations Secretary-General in response to Resolution [A/RES/79/239](#) “Artificial intelligence in the military domain and its implications for international peace and security.”

Founded in 2009, ICRAC is a civil society organization of experts in artificial intelligence, robotics, philosophy, international relations, human security, arms control, and international law. We are deeply concerned about the pressing dangers posed by AI in the military domain. As members of the Stop Killer Robots Campaign, ICRAC fully endorses their submission to this report, and wishes to provide further detail regarding the concerns raised by AI-enabled targeting.

Increasing investments in AI-based systems for military applications, specifically AI-enabled targeting, present new threats to peace and security and underscore the urgent need for effective governance. ICRAC identifies the following concerns in the case of AI-enabled targeting:

1. AI-enabled targeting systems are only as valid as the data and models that inform them. ‘Training’ data for targeting requires the classification of persons and associated objects (buildings, vehicles) or ‘patterns of life’ (activities) based on digital traces coded according to vaguely specified categories of threat, e.g. ‘operatives’ or ‘affiliates’ of groups designated as combatants. Often the boundary of the target group is itself poorly defined. Although this casts into question the validity of input data and associated models, there is little accountability and no transparency regarding the bases for target nominations or for target identification. AI-enabled systems thus threaten to undermine the Principle of Distinction, even as they claim to provide greater accuracy.

2. Human Rights Watch research indicates that in the case of IDF operations in Gaza, AI-enabled targeting tools rely on ongoing and systematic Israeli surveillance of all Palestinian residents of Gaza, including with data collected prior to the current hostilities in a manner that is incompatible with international human rights law.

3. The increasing reliance on profiling required by AI-enabled targeting furthers a shift from the recognition of persons and objects identified as legitimate targets by their observable disposition as an imminent military threat, to the ‘discovery’ of threats through mass surveillance, based on statistical speculation, suspicion and guilt by association.

4. The questionable reliability of prediction based on historical data when applied to dynamically unfolding situations in conflict raises further questions regarding the validity and legality of AI-enabled targeting.

5. The use of AI-enabled targeting to accelerate the scale and speed of target generation further undermines processes for validation of the output of targeting systems by humans, while greatly amplifying the potential for direct and collateral

civil harm, as well as diminishing the possibilities for de-escalation of conflict through means other than military action.

Justification for the adoption of AI-enabled targeting is based on the premise that acceleration of target generation is necessary for ‘decision-advantage’, but the relation between speed of targeting and effectiveness in overall military success, or longer-term political outcomes, is questionable at best. The ‘need’ for speed that justifies AI-enabled targeting is based on a circular logic, which perpetuates what has become an arms race to accelerate the automation of warfighting. *Accelerating the speed and scale of target generation effectively renders human judgment impossible or, de facto, meaningless.* The risks to peace and security – especially to human life and dignity – are greatest for operations outside of conventional or clearly defined battlespaces. Insofar as the use of AI-enabled targeting is shown to be contrary to international law, the mandate must be to *not* use AI in targeting.

In this regard, ICRAC notes that the above systems present challenges to compliance with various branches of international law such as international humanitarian law (IHL), *jus ad bellum* (UN law on prohibition of use of force), international human rights law (IHRL) and international environmental law. In the context of military AI’s implications for peace and security, *jus ad bellum*, a framework that prohibits aggressive military actions and regulates the conditions under which states may lawfully resort to the use of force, is the most relevant. In the same manner IHRL is important in this context because it is designed to uphold human dignity, equality, and justice – values that form the foundation of peaceful and secure societies.

International Humanitarian Law and Youth Initiative

[11 April 2025]

Artificial intelligence (AI) has gained a universal recognition during the 1950s’. Technological emergence has assisted humans in almost all facets of their lives thereby making work easier and faster. Moreso, the rapid growth of Artificial intelligence in technological field enthraling commercial investors, law makers, defense intellectuals and international competitors can be evidential in theoretical premises of international security. The use of Artificial intelligence (AI) in modern warfare particularly in the In the Middle East and North Africa, Ukraine/Russian armed conflict which has resulted in the killings of thousands of innocent civilians with women and children being the most vulnerable. The emergence of AI is expected to be utilized in improving all sectors in our daily lives However, its Negative application in the military domain continues to create Humanitarian crisis between warring parties making it of regional and international concern. The war in Gaza is one of the deadliest and most destructive war in history with technology playing a central role in enabling mass slaughter and destruction ranging from supplying the dystopian systems used to automate the killings and bombing.¹ Following the October 7 2023, there have been extensive reports evidencing the Israeli occupation forces use of surveillance technology, artificial intelligence, and other digital tool to determine who, what and when to attack in Gaza trip. Thus, this violates the principles of international humanitarian law which emphasize the necessity of distinguishing those in active combat and not² and to take necessary precautions when conducting an attack to minimize civilian harm.

¹ Accessnow. (October 2024) Big Tech and the risk of genocide in Gaza: what are companies doing? Available at <https://www.accessnow.org/gaza-genocide-big-tech/>.

² Article 48 of Additional protocol I of the Geneva convention.

IHLYI in this paper, responding to the request of the UN Secretary-General pursuant to a resolution [A/RES/79/239](#), adopted by the General assembly on 24 December 2024 on Artificial intelligence in the military domain and its implication to international peace and security therefore, it analyzes AI In modern warfare, its implication to international peace and security and the role of technological companies in armed conflict.

Artificial Intelligence in Modern Warfare: A Legal and Humanitarian Perspective

The rules of international humanitarian law do not explicitly address the use of modern technological tools and artificial intelligence (AI) during armed conflicts. However, its core principles – such as distinction, proportionality, and precaution – remain applicable and binding on all parties. These principles require the differentiation between military objectives and civilians, and oblige parties to take all feasible measures to avoid or minimize harm to civilian populations. In recent years, militaries have contracted private companies to develop autonomous weapons systems. However, the armed conflict in Gaza stands out as one of the most prominent cases where commercially developed AI models – originally created in countries like the United States – have been employed in actual combat operations, despite the fact that these systems were not initially designed to make life-or-death decisions.

This shift highlights a troubling rise in the militarization of technology without clear legal or ethical oversight. While some of these tools may enhance operational efficiency, their unregulated use poses serious risks of human rights violations, especially amid a lack of transparency about how these tools function, the origin of the data they rely on, and the accuracy of their outcomes³.

One of the most pressing concerns recently raised is the deployment of digital military tools based on unreliable data or flawed algorithms. Some of these systems depend on mass surveillance of Gaza's⁴ population, including the collection of personal data prior to the outbreak of hostilities. Such practices raise legal and ethical questions regarding their compatibility with international obligations to safeguard privacy and prevent the misuse of personal information for the purpose of direct targeting.

Among the tools reportedly in use is a system that tracks population movement through mobile phone data to monitor evacuations from certain areas. Another generates lists of structural targets to be hit militarily. A third tool classifies individuals based on levels of suspicion regarding their affiliation with armed groups, while a fourth seeks to determine the precise location of a target in order to carry out a strike at the opportune moment. These tools largely rely on data extracted from mobile devices – whether through cell tower location information or GPS⁵. However, from a technical perspective, such data is insufficiently precise to confirm an individual's presence at a specific location at a given time, particularly in conflict zones where individuals frequently change phones or numbers. Over-reliance on this technology may lead to fatal mistakes, especially when a mobile phone is used as a substitute for verifying a person's actual presence in a targeted area. Legally, the use of such systems without taking all feasible precautions to protect civilians constitutes

³ Human Rights Watch, "Israel: AI-Powered Targeting Systems May Be Committing War Crimes in Gaza", 2024.

⁴ Associated Press, "Documents Reveal Israel's Use of AI Tools in Targeting Gaza", Investigative Report, 2024.

⁵ Human Rights Watch (2024). Questions and Answers: Israeli Military's Use of Digital Tools in Gaza Available at Questions and Answers: Israeli Military's Use of Digital Tools in Gaza | Human Rights Watch.

a clear violation of international humanitarian law – particularly Article 57⁶ of Additional Protocol I to the Geneva Conventions, which obliges parties to take constant care to spare civilian lives during military operations.

Given this reality, urgent questions must be raised about the future of AI in warfare and the legislative and legal mechanisms needed to regulate it. Without proper oversight, these tools risk becoming instruments of systematic human rights abuses rather than technologies aimed at ensuring greater protection for those affected by war.

Implications of Artificial Intelligence on International Peace and Security

Armed conflicts in various regions around the world, such as Gaza, Lebanon, Syria, Ukraine, and Libya, have had catastrophic humanitarian and security consequences. These conflicts have led to the mass displacement of civilian populations, depriving thousands of people of their basic rights such as food, water, shelter, and healthcare. These individuals live in dire humanitarian conditions, with a significant increase in deaths due to famine, thirst, and diseases caused by contaminated water, in addition to exposure to harsh weather conditions without protection.

In this context, the increasing use of artificial intelligence and drones as weapons in conflicts, particularly by Israel in the Gaza Strip⁷, stands out. Since October 2023, there has been a notable escalation in the use of “quadcopters” to carry out precise and targeted strikes against civilians. These drones are equipped with data analysis algorithms and offensive capabilities, enabling them to target individuals based on tracking their movements or mobile phone signals.

According to documented reports, this technology has led to the death of more than 1,000 Palestinians by May 2024, including a significant number of women and children. This constitutes a grave violation of international humanitarian law, particularly Articles 51 and 57 of Additional Protocol I to the Geneva Conventions, which prohibit attacks on civilians and oblige parties to the conflict to take all necessary precautions to avoid harming them.

The concerns are not limited to the use of artificial intelligence against individuals but extend to the misuse of data. Relying on mobile phone tracking technologies (either through GPS data or cell tower signals) to pinpoint individuals’ locations presents serious risks. Recent studies have shown that these systems do not provide enough accuracy to reliably determine someone’s location, especially in conflict zones where phones may be swapped or disconnected frequently. This means that relying on these methods without field verification can lead to erroneous decisions, resulting in unlawful killings.

In a well-known case, a Palestinian woman named “Silah” was killed while carrying a white flag and leading her family to safety. After stepping onto a main street, she was targeted by a small drone that shot her in the head. This incident, witnessed by those around her, serves as a stark example of the disastrous outcomes of unregulated use of technology on the battlefield⁸.

In Libya, drones played a decisive role in the battles between conflicting parties, particularly as many of these drones, including Turkish and Chinese models, were operated using data analysis systems to target objectives. Some of these systems are

⁶ Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), Article 57.

⁷ TRTWORLD (2024) Quadcopter strikes: 1000 Palestinians killed by Israeli drones in one year. Available at Quadcopter strikes: 1000 Palestinians killed with drones in a year.

⁸ Gaza grandmother gunned down by Israeli sniper as child waved white flag,” *Times Kuwait*, November 2024, <https://timeskuwait.com/news/gaza-grandmother-gunned-down-by-israeli-sniper-as-child-waved-white-flag>.

believed to rely on artificial intelligence techniques for targeting, without legal oversight. The use of these tools in urban areas like Tripoli and Sirte has led to the deaths of civilians and extensive damage to infrastructure⁹.

All of these events indicate that integrating artificial intelligence into managing and directing armed conflicts without an internationally binding legal framework to regulate its use could open the door to widespread violations, especially if these systems are not subject to independent and transparent oversight to ensure compliance with international humanitarian law and human rights.

Roles of Companies Developing AI in Armed Conflicts

Through a rapid increase in artificial intelligence and computer services, U.S. tech corporations have discreetly given Israel the ability to monitor and kill many more militants in Gaza and Lebanon more quickly. However, the death toll among civilians has also skyrocketed, raising concerns that these instruments may be causing the deaths of innocent people. Israel's recent wars are a leading example of commercial AI models developed in the United States being used in active warfare, despite concerns that they were not originally designed to help decide who lives and who dies.

For years, militaries have hired private companies to create customized autonomous weapons. Numerous American software companies have backed Israel's battles in recent years, including Microsoft and the San Francisco-based startup OpenAI. Under "Project Nimbus," a \$1.2 billion contract signed in 2021¹⁰ when Israel first tried out its in-house AI-powered targeting systems, Google and Amazon offer cloud computing and artificial intelligence services to the Israeli military. The military has made use of Dell and Cisco data centers and server farms. Palantir Technologies, a Microsoft partner in U.S. defense contracts, has a "strategic partnership" that provides AI systems to support Israel's war efforts, while Red Hat, an independent IBM company, has also supplied cloud computing technologies to the Israeli military.

Furthermore, through a number of programs, Microsoft also supplies Israel's government with services that have allegedly been used to help the Israeli military, police, Israeli Prison Service (IPS), and illegal settlement operations. Over 10,000 Palestinians are being held by the IPS as of October 2024; half of them have been detained without being charged or having a trial date scheduled. At least 310 medical professionals, UN employees, women, and children are among the Palestinian prisoners from Gaza who are presently detained in prolonged, secret, and incommunicado detention, where they are subjected to torture, mistreatment, and sexual violence and abuse, according to the UN Human Rights Office.

Companies are under obligation to respect human rights within their scope of operations. Companies that directly aid the offender – for example, by offering financial, logistical, military, or intelligence support – may be held criminally responsible for a crime committed during an armed conflict. Companies and their managers or executives may be held accountable in certain situations even if they had no direct involvement in the crime or no intention of supporting it. As the Office of the High Commissioner on Human Rights (OHCHR) noted, companies "should treat

⁹ France 24. (2021). "Have Killer Drones Been Deployed in Libya?". France 24. Retrieved from <https://rb.gy/1m6k43>.

¹⁰ APNEWS (2025). As Israel uses US-made AI models in war, concerns arise about tech's role in who lives and who dies. Available at How US tech giants' AI is changing the face of warfare in Gaza and Lebanon | AP News.

this risk in the same manner as the risk of involvement in a serious crime, whether or not it is clear that they would be held legally liable¹¹.”

In light of the concerns raised in this submission and their implications for international peace and security, IHLYI urges states to:

1. Refrain from the use of AI in military applications: States should immediately halt the use of artificial intelligence in military activities and establish national regulations and laws to prevent its deployment in warfare.

2. Work towards a global ban on the military use of AI: States should actively pursue international agreements and frameworks to ban the use of AI in military contexts, ensuring that no country utilizes AI for warfare.

3. Avoid the development of autonomous and AI-enabled weapon systems: States should refrain from developing autonomous weapon systems or AI-powered weaponry that could be used to target humans, ensuring human oversight and decision-making in military actions.

4. Ensure the protection of personal data: States must guarantee that personal data is protected from misuse by military forces, law enforcement agencies, border control, and private contractors collaborating with these entities.

5. Promote accountability in AI development: Technology companies, researchers, engineers, and financial institutions should commit to not supporting the development or funding of AI technologies designed for military applications, advocating for responsible innovation in line with humanitarian principles.

Peace Movement Aotearoa and Stop Killer Robots Aotearoa New Zealand

[21 May 2024]

Peace Movement Aotearoa and Stop Killer Robots Aotearoa New Zealand welcome the opportunity to contribute our views to the UN Secretary-General’s report on artificial intelligence (AI) in the military domain and its implications for international peace and security. Our submission briefly outlines our involvement in this issue, and has three sections summarising our position on: a) A new international instrument on military use of AI and autonomy in weapon systems is urgently needed; b) Key focuses of a new international instrument; and c) Scope of a new international instrument. The points below are based on discussions with our member and supporting groups about the content of this submission.

Introduction

Peace Movement Aotearoa is the national networking peace organisation in Aotearoa New Zealand, established in 1981 and registered as an Incorporated Society in 1982. Our purpose is networking and providing information and resources on peace, humanitarian disarmament, human rights and social issues; and we have extensive national networks of member and supporting groups and individuals. We are a founding member of the Stop Killer Robots campaign and coordinate the national Stop Killer Robots Aotearoa New Zealand (SKRANZ) campaign.

SKRANZ was launched in April 2013 to support the global campaign, with a specific national focus on urging New Zealand to take national action to prohibit the development, production and use of autonomous weapon systems; and to take

¹¹ Accessnow (2024) Big Tech and the risk of genocide in Gaza: what are companies doing?
<https://www.accessnow.org/gaza-genocide-big-tech/>.

international action to support negotiations on a new treaty to prohibit autonomy in weapon systems. Since 2023 we have widened our focus to include military use of AI as its perils became increasingly obvious.

(a) A new international instrument on military use of AI and autonomy in weapon systems is urgently needed

As outlined in our submission for the UN Secretary-General's report on autonomous weapon systems ([A/RES/78/241](#)) last year, it has been clear for some years now that rapidly developing technological advances in the use of force and increasing autonomy in weapon systems pose an unprecedented threat both to humanity and to the foundations of international human rights and humanitarian law, which are based on respect for human life and dignity, protection of humanity in times of oppression and armed conflict, and human responsibility and accountability for harm.

The serious ethical, humanitarian, legal, and security concerns posed by these developments have been discussed for more than a decade within United Nations bodies – including the Human Rights Council, meetings related to the Convention on Certain Conventional Weapons and in the UN General Assembly – as well as in regional and national governmental and non-governmental forums.

Even as these discussions have taken place, some states have increasingly incorporated autonomy into military use of force in ways that have already resulted in gross violations of international law with disastrous consequences for civilian populations. It is apparent that the absence of specific international law on autonomy in weapon systems, and with differing interpretation by some states as to how existing law applies to new technological developments, the risk of proliferation of ever more dangerous and uncontrollable weapon systems is increasing rapidly.

The need for urgency for international action on this has been highlighted over the past eighteen months by, for example, Israel's use of AI-powered target suggestion systems in Gaza to make high explosive strikes on numerous targets possible in a short time frame, resulting in indiscriminate slaughter of civilians and systematic destruction of life-sustaining infrastructure. The reality of digital dehumanisation with catastrophic consequences is now very evident, as is the increasing tendency towards the development and use of autonomous weapon systems that will remove any remaining vestige of humanity from war.

We have noted with concern that states who brought forward [A/RES/79/239](#) include states that have armed and supported Israel's genocidal attacks on Gaza, and where big data tech companies contributing data storage and AI capabilities to Israel's military systems are based.

Similarly, 'responsible AI in the military domain' (surely an oxymoron) is being promoted by states already developing their own AI targeting and autonomous weapon systems, as a way of undermining the push towards a binding instrument to prohibit these critical threats to international peace and security.

The US 'Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy' has highlighted for us the risk of horizontal proliferation of both military use of AI and autonomous weapon systems as states that do not have their own capability in this regard move from interoperability to integration with the states of armed forces that do or that are developing it. In the case of New Zealand, for example – as it seeks to be a 'combat capable force multiplier with enhanced lethality'¹² – this involves closer military integration with the armed forces of

¹² See, for example, the 2025 Defence Capability Plan released this week.

Australia and the US: New Zealand endorsed the US ‘Political Declaration’ early last year specifically to be compliant with US military doctrine.

These unfortunate developments underscore the urgent need for a new international instrument on military AI and autonomy in weapon systems to clarify and strengthen existing law. The instrument must include both prohibitions and regulations, as outlined below, and must include military use of AI in combat.

As emphasised in the UN Secretary-General’s 2024 report on autonomous weapon systems¹³, negotiations on a new instrument must begin without any further delay, in a multilateral forum where states can come together to work constructively, where the voices of those whose lives have already been impacted by military use of AI and increasing autonomy in weapon systems can be heard, and where UN agencies, the International Committee of the Red Cross (ICRC), and NGOs are active participants.

(b) Key focuses of a new international instrument

While much of the work around military use of AI and autonomous weapon systems has focused on the issue of meaningful human control over the use of force, it is our view that the key underlying ethical imperative is preventing human beings from being targeted or attacked by any system utilising digital code and/or sensors. A prohibition on military use of AI and autonomy in weapons systems that are designed or used to target human beings must be the starting point.

Meaningful human control over the use of force clearly has an ethical component, but it is also a practical and legal means to ensure accountability for any autonomy in weapon systems that breach the key dictates of humanitarian law.

(c) Scope of a new international instrument

It is our view that a new international instrument should include overarching rules to establish a framework for evaluating current and future technological developments, while promoting increased compliance with international human rights and humanitarian law.

Such overarching rules would prohibit autonomous weapon systems that are designed or used to target humans, and lay out specific obligations to ensure meaningful human control over other systems: for example, that the human operator/s understand the capabilities and limitations of the system, are able to fully evaluate the context in which the system will be used, and are making mindful firing decisions rather than assuming the technology is accurate – this would act to regulate autonomy in weapon systems. It would be useful to specify that decisions made by states on their assessment of new or altered weapon systems that incorporate autonomous features or functions must be transparent.

Furthermore, in the context of the UN Secretary-General’s forthcoming report on AI in the military domain and in the light of the awful consequences of military use of AI in Gaza, the scope of a new international instrument must go beyond autonomous weapon systems. It is very clear that there is a spectrum of harmful military use of autonomy, ranging from target decision support systems (as some have described systems such as Lavender), data-based targeting systems, generation of target lists by algorithm or AI, sensor-based targeting systems, through to weapon systems that combine these elements and incorporate varying degrees of machine learning to make target selection decisions and attack autonomously.

We note the 2023 Joint Call by the UN Secretary-General and ICRC President stated “*The autonomous targeting of humans by machines is a moral line that we*

¹³ Lethal autonomous weapons systems: Report of the Secretary-General ([A/79/88](#)), 1 July 2024.

*must not cross*¹⁴, yet that has already happened – a point reiterated in the UN Secretary-General’s 2024 report¹⁵.

It is therefore our view that a new instrument must cover military use of AI – including systems that automate significant decision-making in the use of force, such as target generation, force deployment, and engagement – as well as autonomous weapon systems.

Finally, although we have referred in this submission to military use of AI and autonomy in weapon systems, prohibitions and regulations in a new international instrument must also apply to all coercive agencies of the state, including those used for policing and internal security, for border control, in corrections facilities and in places of detention.

Ploughshares

[11 April 2024]

Project Ploughshares, a Canadian peace research institute, has for over a decade focused its advocacy and research on the military applications of emerging technologies, including artificial intelligence (AI) and autonomous weapons. As AI systems are rapidly advancing and being tested in contemporary conflict zones, international governance frameworks have struggled to keep pace. Meanwhile, intensifying geopolitical competition increases the likelihood that AI technologies will be deployed in complex, dynamic environments for which they are not suited – raising significant risks for civilians.

The wide-ranging use of AI in military applications demands urgent and coordinated international attention. We encourage the Secretary-General and member states to focus on three particularly pressing areas: the use of AI in decision-support systems related to the use of force, the dual-use nature of AI technologies, and the widening capacity gap among states engaging in multilateral discussions.

AI decision-support systems

One area that remains insufficiently addressed in current international discussions is the use of AI in military decision-making, especially decisions about the use of force. Of particular concern are AI-enabled targeting tools such as “Lavender” and “Gospel,” reportedly used in Gaza. These systems are classified as “decision support” because a human is technically required to approve target selections. However, there is little transparency regarding how these decisions are made, how frequently AI-generated recommendations are rejected, or whether human operators fully understand how the AI systems reach their conclusions.

In practice, these systems raise the risk of “rubber-stamping,” in which human oversight becomes superficial, thereby undermining the principle of meaningful human control and increasing the likelihood of harm to civilians. The potential use of such AI systems in early-warning, surveillance, reconnaissance, and nuclear command-and-control systems further amplifies these concerns.

To mitigate these risks, states must work toward clear norms, regulations, and training requirements that enhance operator understanding, counter automation bias, and ensure genuine human engagement in decision-making processes.

¹⁴ Joint call by the United Nations Secretary-General and the President of the International Committee of the Red Cross for States to establish new prohibitions and restrictions on Autonomous Weapon Systems, 5 October 2023.

¹⁵ As at note 3.

Dual-use challenges

AI's dual-use nature – its applicability to both civilian and military domains – creates further governance complexity. Civilian-developed technologies can be repurposed for military use without appropriate testing or safeguards, increasing the risk of conflict escalation, misuse, and error. Additionally, the accessibility of certain AI tools means that nonstate armed groups may also gain access, potentially using them to target civilians and infrastructure.

We urge states to develop policy mechanisms, including export controls, technology impact assessments, and multistakeholder engagement, to account for dual-use risks and promote responsible innovation.

Capacity- and knowledge-building

Current multilateral discussions reveal stark capacity disparities among states, many of which do not have the resources or technical expertise to participate meaningfully in governance efforts. To ensure inclusive and equitable global engagement, we recommend that states collaborate with the UN Office for Disarmament Affairs to strengthen capacity-building initiatives.

The scientific and academic communities also have a role to play in supporting the development of accessible resources and training materials. International forums, such as the upcoming REAIM Summit in Spain, should include dedicated sessions for knowledge-sharing, especially to support representatives from under-resourced states.

Final thoughts

The international community is at a crossroads. The accelerating militarization of AI demands robust diplomatic responses. We can – and must – move from aspirational principles to concrete, enforceable frameworks, by employing political will, inclusive dialogue, and cross-sector collaboration.

AI-powered warfare is no longer a theoretical risk; it is a present reality. Whether this new era enhances global security or undermines it will depend on the steps states take now to strengthen governance, manage technological competition, and uphold international humanitarian norms.

Without timely, coordinated action, the risks of accidental escalation and unintended conflict will only increase.

Soka Gakkai International

[10 April 2025]

The Soka Gakkai International (SGI) welcomes the opportunity to share our views on the important issue of artificial intelligence (AI) in the military domain. As an NGO whose work is guided by Buddhist principles, we urge that the United Nations, its Member States and other stakeholders take into careful consideration the impact of AI in the military domain from a standpoint of upholding and respecting human dignity.

Introduction

AI in the military domain is rapidly evolving and transforming modern warfare and international peace and security. These systems are being used for various purposes, including surveillance, autonomous weapons, decision-making support, and logistics. With such wide-ranging applications, the integration of AI technologies in military systems poses significant challenges. To better ensure compliance with

international humanitarian law (IHL) and uphold protection for civilians and combatants alike there are several issues that we may consider.

Lack of transparency and accountability

- If an AI system were to make an error – such as identifying a target incorrectly – it could be difficult to pinpoint the cause of the error, “the black box problem”. Was it a flaw in the data used to train the AI, an issue with the algorithm, or a problem in the operational context? Without transparency within these systems, assigning responsibility is difficult.
- International laws and treaties, such as the Geneva Conventions, were created before AI systems became commonplace in warfare. Without global norms and legal frameworks, there is no consistent approach to ensuring accountability for AI decisions made in warfare.
- With inadequate accountability mechanisms in place, AI could be used for military strategies that violate human rights, suppress civil liberties, or engage in unethical operations.

Speed of decision-making and risk of escalation

- The ability of a military force to make decisions and execute actions faster than its opponent is increasingly viewed as having a strategic advantage. However, the drive for speed can lead to unintended and costly consequences.
- Decisions made too quickly without proper analysis or consideration can lead to poor outcomes, including tactical blunders, strategic missteps, or ethical violations.
- Instead of diffusing a tense situation or negotiating, if combatants react too quickly it could provoke an even greater confrontation, further escalation and prolonged conflict resulting in more human suffering including amongst civilians.
- The acceleration of decision-making processes closes down the possibility of meaningful human control, the growing trend to automate decision-making threatens the ability to achieve human oversight which is essential to facilitate compliance with IHL.

Bias in AI in the military domain

- AI bias refers to the presence of systematic and unfair discrimination in AI systems, such as historical bias, where systems may reinforce harmful stereotypes, bias in data processing and algorithm development which can lead to making biased decisions and bias in how the systems are used.
- AI bias in the military domain is a significant concern, particularly as AI systems are increasingly being integrated into defense and security operations. The potential for AI bias to emerge in these areas can result in human rights implications, exacerbating existing inequalities and lead to deadly consequences for certain groups.
- AI heavily relies on vast amounts of high-quality and reliable data for decision-making. There are several potential violations when it comes to obtaining this data including issues around privacy and surveillance, challenges of bias also arise when dealing with incomplete and inaccurate data.

- When AI systems are biased, they not only perpetuate inequalities but also contribute to the digital dehumanization¹⁶ of marginalized groups.

Proliferation

- Nations may rush to develop AI-based military technologies to outpace their adversaries, which could lead to a destabilizing arms race and increased global tensions.
- Without regulation autonomous weapons systems in particular, could proliferate globally, including amongst non-state actors which could increase crime nationally and regionally, exacerbating social inequalities, overwhelm resources and infrastructures of countries, as well as undermine social and national security.

Conclusion

The issue of AI within the military contexts is complex, and without regulation, it could lead to serious consequences for global peace and security. The desire to speed up decision-making processes within this context has yet to be proven as an effective way of resolving conflicts and achieving peace and security. Furthermore, you cannot divorce AI in the military and AI in civil uses, a failure to address AI in a military context could have widespread repercussions in all spheres of civil life including law enforcement, border control, education, housing and health care. Fundamentally, AI is here to stay, how we utilize it in the military and in our lives will shape the course of humanity. We have the possibility and the responsibility to decide how we want to use technology, knowledge, and the world's resources. To use it in a way that uplifts humanity or degrades it? This is an urgent question that requires moral, ethical and courageous leadership.

Stop Killer Robots

[11 April 2025]

The Stop Killer Robots campaign welcomes the opportunity to submit our views to the United Nations Secretary-General in response to Resolution [A/RES/79/239](#).

Established in 2012, we are a coalition of more than 270 non-governmental organisations working across 70 countries.¹ We seek to counter threats to humanity and human dignity through the adoption of a new international treaty to prohibit and regulate autonomous weapons systems.² We support the development of legal and other norms that ensure meaningful human control over the use of force, counter digital dehumanisation, and reduce automated harm.³

Building an effective international response to emerging technologies

Autonomous weapons systems, 'AI in the military domain,' and trends and developments in increasingly automated decision-making and action in the use of force – as well as in our lives and societies more broadly – are all part of the same concerning picture:

The growing influence of computer processing and algorithmic thinking increasingly shapes our interactions in the world and the outcomes available to us.

¹⁶ Digital dehumanization is a process where humans are reduced to data, which is then used to make decisions and/or take actions that negatively affects their lives.

¹ See www.stopkillerrobots.org/about-us and www.stopkillerrobots.org/a-global-push/member-organisations.

² See <https://www.stopkillerrobots.org/our-policies/>.

³ See www.stopkillerrobots.org/vision-and-values/.

There are clear threats to peace, justice, dignity, human rights, equality, responsibility and accountability, and respect for law. We are getting closer to machine processes determining whom to kill.

To address these challenges effectively, a comprehensive and holistic response is needed from the international community.

Adopting a legally binding instrument on autonomous weapons systems will be one critical component: we must draw basic red lines for humanity against the automation of killing, which brings under jeopardy both international humanitarian law and international human rights law, in particular the presumption of innocence, the right to equality and non-discrimination, dignity, and wipes away contextual circumstances of the target(s) in question. The UN Secretary-General's comprehensive report last year reiterated his urgent call on states to negotiate a legally binding instrument to prohibit and regulate these systems by 2026.

But, a new international treaty on autonomous weapons systems alone may not be enough. States must also reach agreement on preventing and addressing grave harm from other uses of emerging technologies. A whole set of strong international rules are needed that stop the erosion of meaningful human control and the slide towards greater digital dehumanisation and automated harm, across international and domestic practice, in armed conflict and in civilian life.

'Military applications of AI' are already contributing to civilian harm

The risks of integrating AI into the use of force in armed conflict reach far beyond those to peace and security between states: a holistic consideration of peace and security that considers dimensions such as ethical, legal, and humanitarian issues must be taken into account in the UN Secretary-General's report under resolution [79/239](#).

We are already seeing grave threats to civilian protection and human rights and huge harm being caused by AI and automation in the use of force. This is arising from the quest for speed in warfare, the reduction of people to objects, and issues such as automation bias that Stop Killer Robots has raised the alarm about for years.

We have been horrified by reports of the use of AI-powered 'decision support systems' by Israel in Gaza, which suggest human targets to strike.⁴ According to reports, human approval of these suggestions in vast volumes at high speed has been minimal – entailing digital dehumanisation, the erosion of meaningful human decision-making and control (including through automation bias), and directly contributing to massive and devastating harm to civilians in Gaza, alongside other tools.⁵

Many other states are developing and using such 'decision support systems', which raise concerns around international humanitarian law, human rights law, and ethics. So far there are few reports on how these are being deployed, with what constraints and with what impacts. Nevertheless, the push by many states to develop and integrate AI and autonomy into decision-making and the use of force is a huge concern. The further use in hostilities of these kinds of tools by any state in the unacceptable ways that we have seen in Gaza must be prevented. Stop Killer Robots struggles to see how such uses could meet the definition of the responsible application of AI in the military domain given in resolution [79/239](#).

⁴ 'Lavender': The AI machine directing Israel's bombing spree in Gaza, +972 Magazine <https://www.972mag.com/lavender-ai-israeli-army-gaza/>.

⁵ Questions and Answers: Israeli Military's Use of Digital Tools in Gaza, Human Rights Watch, <https://www.hrw.org/news/2024/09/10/questions-and-answers-israeli-militarys-use-digital-tools-gaza>.

Further risks to peace and security, rights, and human dignity

The quest for greater speed through AI and automation – towards the goal of increasing the tempo of conflict to a point beyond human cognition in the pursuit of a military and strategic edge – is an extremely dangerous one for international peace and security. These risks are further to the impact ‘AI in the military domain’ is already having on civilian protection. Risks include unwanted escalation, lowered political thresholds to the use of force, and arms race dynamics.

Technologies that can contribute to target selection (such as threat detection tools) and remote biometric surveillance (such as facial recognition) have already had documented negative impacts on human rights such as the rights to privacy, equality and non-discrimination, freedom of expression and peaceful assembly, and the freedom of movement. In the case of facial recognition for identification (1:n), the technology is considered by many legal experts as wholly incompatible with international human rights law.

That AI systems inevitably encode and reproduce the biases of our societies – including racism, sexism and ableism – and that such bias cannot be eliminated, is also well established. The use of such systems to process people in the use of force will inevitably lead to disproportionate – and multiplied – impacts on already marginalised and minoritised people. Integrating automation and AI into decisions and actions in the use of force against people contributes to digital dehumanisation – the process where humans are reduced to data, which is then used to make decisions and/or take actions that negatively affects their lives.

The relationship with autonomous weapons systems

Stop Killer Robots notes that the UN Secretary-General’s report will be on the “application of artificial intelligence in the military domain, with specific focus on areas other than lethal autonomous weapons systems.” It is important nevertheless to highlight that various applications beyond the boundary of autonomous weapons systems are closely linked to them.

Firstly, such tools could be integrated as components of autonomous weapons systems now or in the future. For example, a ‘decision support system’ could be used as an autonomous targeting system, connected to a platform tasked to strike targets on the list generated, based on processing sensor data. Secondly, these tools are linked not only practically, but raise and are part of the same picture of concern. Strikes undertaken based on the nominal human approval of targets generated by a decision support system do not sit far from strikes undertaken with an autonomous weapons system.

It is therefore important that states consider these issues in dialogue: many of the rules and principles developed for autonomous weapons systems on keeping control and rejecting automated killing will need to be extended (with adaptations) to other tools; and, how the development of AI in the military domain more broadly will impact the direction and challenges posed by autonomous weapons systems will need consideration.

Recommendations

Technologies incorporating AI and automation into the use of force in armed conflict are currently being deployed without specific agreed rules; the principles various states have proposed and committed to so far have been too weak and vague to prevent civilian harm and risks to peace and security.

All developments in autonomy and AI in the use of force which threaten our safety, security, and humanity must be urgently and adequately addressed through strong regulation by the international community, with unacceptable uses prevented.

States must:

- Move with urgency to negotiate and adopt a new international treaty to prohibit and regulate autonomous weapons systems;
- In International discussions, critically and meaningfully engage with the implications and real-world consequences of current practice in the use of tools that fall under ‘AI in the military domain,’ including acknowledging and examining humanitarian harm;
- Fully consider the legal, ethical, humanitarian, and peace and security risks of further development and use of such systems, whatever the perceived ‘benefits’ may be
- Work urgently to prevent unacceptable uses of technology and trends in development, through committing to develop strong norms for meaningful human control and against digital dehumanisation:
 - This should take place domestically, regionally, and internationally.
 - It must involve a comprehensive and holistic international response, including a legally binding instrument prohibiting and regulating autonomous weapons systems alongside other measures.
 - It should include consideration and development of the other legal instruments necessary to preserve meaningful human control and to protect human dignity against AI in the use of force.

Stop Killer Robots Youth Network

[10 April 2025]

The Stop Killer Robots Youth Network welcomes the opportunity to submit recommendations for consideration by the United Nations Secretary-General in response to Resolution [79/239](#) “Artificial intelligence in the military domain and its implications for international peace and security” adopted by the General Assembly on 24 December 2024. As a global network of young people under age 30 in over 50 countries working to secure a future free of automated killing, we have consistently advocated for the creation of a new treaty on autonomous weapons systems (AWS) – in particular, we insist on a total prohibition of anti-personnel autonomous weapons as we wish to build a world without such dehumanising weapons. While youth will inevitably face the risks of new weapons technologies, we remain underrepresented in the decision-making process and are often sidelined in forums that shape our interests. As youth who have grown up in an increasingly digital world, we wish to create a future where technology is used to promote peace, justice, equality, and human rights, not perpetuate violence.

With escalating conflicts and the rapid deployment of new weapons technologies around the world, there is an urgent need to reinvest in international law as a measure to build trust and achieve sustainable peace and security. The application of artificial intelligence (AI) in the military domain presents numerous challenges that concern us as youth, including digital dehumanisation, the gamification of violence, and the further erosion of human control and involvement over the use of force.

Military AI & AI systems already in use

Artificial intelligence has been progressively implemented in the military domain over the past decade, however, due to the opacity of military activities and development, the wide public has not been aware of this issue until recently when the active uses of AI systems have been mediated. We have seen and monitored the use

of AI systems to support the targeting of both objects and people. Unfortunately, the use of such systems have not been able to alleviate civilian suffering, for example, in Gaza where one third of victims are children and where too many civilian infrastructures, including critical infrastructures such as humanitarian camps, hospitals¹, and schools², have been either directly targeted or indirectly impacted by the hostilities.

There have been other concerning uses³ of AI systems outside of the military which need to be considered as they might be implemented in the military domain, mainly predictive AI and facial recognition. Predictive AI technologies have been used in the police and judicial systems since the early 2010s and have been shown to be ineffective, incorrect, and subject to reinforcing discriminatory behavior.⁴ If predictive AI were to be implemented in the military domain, it could lead to the increasing risk of civilians being targeted as they could be labeled as possible fighters or being indirect victims of military activities due to the multiplications of targets with predicted military advantages. Facial recognition technologies (FRTs) are also of concern as they are also unreliable especially when it comes to the identification of non-white males. Facial recognition-enabled targeting in military operations must be prohibited as those systems cannot comprehensively analyse every factor that makes military personnel or civilians a target or not.

Digital dehumanisation

One of the main concerns we have about the use of AI systems in the military domain is the proliferation and banalisation of “**Digital dehumanisation**”. We define digital dehumanisation as the process whereby humans are reduced to data, which is then used to make decisions and/or take actions that negatively affect their lives. This process deprives people of dignity, demeans individuals’ humanity, and removes or replaces human involvement or responsibility through the use of automated decision-making in technology.⁵ Additionally, the increased speed and scale of target production through military AI erodes moral restraints in war and lowers the impact and capacity of decisions from human operators⁶, thus enabling the AI systems to make decisions without meaningful human control, which further dehumanises the decision-making process.

¹ World Health Organization (2025), ‘oPt Emergency Situation Update’. https://www.emro.who.int/images/stories/Sitreps_57.pdf.

² Save the Children (2025), ‘Education Under Attack In Gaza, With Nearly 90% Of School Buildings Damaged Or Destroyed’. <https://www.savethechildren.net/blog/education-under-attack-gaza-nearly-90-school-buildings-damaged-or-destroyed>.

³ Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica (2016), ‘Machine Bias’. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

⁴ Will Douglas Heaven, MIT Technology Review (2020), ‘Predictive policing algorithms are racist. They need to be dismantled’. <https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/>.

⁵ Automated Decision Research (2022), ‘Autonomous weapons and digital dehumanisation’. <https://automatedresearch.org/news/report/autonomous-weapons-and-digital-dehumanisation-a-short-explainer-paper/>.

⁶ Marta Bo and Jessica Dorsey, OpinioJuris (2024), ‘Symposium on Military AI and the Law of Armed Conflict: The ‘Need’ for Speed – The Cost of Unregulated AI Decision-Support Systems to Civilians’. <http://opiniojuris.org/2024/04/04/symposium-on-military-ai-and-the-law-of-armed-conflict-the-need-for-speed-the-cost-of-unregulated-ai-decision-support-systems-to-civilians/>.

Relying on (Big) data leads to problems

We also believe that the use of (big) data in the military leads to multiple issues which need to be considered.

One of the primary issues is the challenge of data labeling – the process of categorizing and tagging data to train algorithms. Inaccurate or biased labeling can have far-reaching consequences, particularly in the context of distinguishing between combatants and non-combatants in conflict zones. If the data used to train military AI systems is flawed or biased, it can lead to disastrous mistakes, such as the targeting of innocent civilians or misidentification of threats.

A critical issue when relying on big data is that the nature data itself is often broken and is incomplete. This means that the data used to train AI models can be incomplete, outdated, or unrepresentative of real-world situations. Such flaws in data can lead to systems that fail to generalize properly, resulting in inaccurate or incorrect predictions and decisions. For example, in combat situations, a lack of diversity in the data used to identify individuals could lead to inaccurate targeting, with devastating consequences. Important data might be missing or poorly represented, such as the exact location of civilians or combatants, which can lead to AI failing to make informed and balanced decisions. In a war scenario, a system trained with data from a specific past conflict may not be capable of handling a new, unpredictable situation. For instance, an AI system that has been fed data from one particular type of conflict might struggle to apply that data to a war with entirely different characteristics, resulting in errors in target identification or incorrect decision-making.

Another significant problem is that many AI systems operate as black boxes. This means that while these systems make decisions and predictions based on the data they process, the decision-making process is not transparent or easily understood. In military scenarios, where the consequences of decisions are extremely serious, the lack of transparency is particularly concerning. If an AI system makes an error, such as wrongly identifying a civilian as a combatant, the absence of clarity about how the system reached that conclusion makes it nearly impossible to understand the origin of the error. This makes accountability difficult, as we cannot determine why the system acted in a particular way. The lack of explanation regarding the decision-making processes of AI also makes it impossible to correct or adjust the system's behavior, potentially perpetuating errors without the ability to fix them effectively.

Linguistic and cultural bias embedded in data which is used to train AI systems can create security vulnerabilities and catastrophically misinterpret communications, behaviors, and intentions across diverse cultural contexts, potentially triggering lethal automated responses to misunderstood signals.⁷ These systems risk automating and amplifying existing prejudices at unprecedented scale and speed with life or death consequences in conflict zones where cultural misunderstandings could rapidly escalate into devastating military actions causing dire consequences.

Accountability

The inclusion of AI systems in the command and decision-making chains will indubitably lead to a lack of accountability and liability for those relying on these systems to make decisions. It will create a sense of distance and lack of liability on the consequences of a decision which mean that decisions may be made without specific, consistent and thorough analysis of the lawfulness and humane characters of

⁷ Jimena Sofia Viveros Álvarez, Humanitarian Law & Policy (2024), ‘The risks and inefficiencies of AI systems in military targeting support’. <https://blogs.icrc.org/law-and-policy/2024/09/04/the-risks-and-inefficiencies-of-ai-systems-in-military-targeting-support/>.

the decision. Then, if an action taken using AI systems violates IHL, the people involved in the implementation and those involved in the decision-making should be held accountable and the use of an AI system shall never exempt people from their responsibilities.

We recognize that military operations are bound by multiple bodies of law – national law, International Humanitarian Law (IHL) and International Human Rights Law (IHRL) – which need to be respected and implemented in order for operations to be lawful. Unfortunately, rules of engagement and of targeting – and all the exceptions – cannot be fully understood and implemented by AI systems. Concepts like doubt, proportionality, and the balance between humanity and necessity are inherently human judgments that cannot be captured by an algorithm. Machines cannot be trusted to uphold these standards on their own. Therefore, it is critical that AI systems never act in a vacuum and that humans retain oversight and decision-making power at all times.

What the future might look like

While AI theoretically has the potential to enhance precision and efficiency in military operations, its integration into warfare raises significant concerns about the future of global security. Autonomous weapons systems, capable of making life-or-death decisions without human control, introduce ethical dilemmas and risks of unintended consequences. The use of AI in military technology is likely to aggravate the existing arms race, as nations compete to develop increasingly sophisticated AI systems, widening the power gap between technologically advanced countries and those less developed, leaving them vulnerable in terms of military readiness. The deployment of autonomous weapon systems and AI-driven tools makes conflict more unpredictable, scalable, and asymmetric, granting certain nations the ability to unleash devastating technologies that smaller states or non-state actors may not be able to counter. The proliferation of AI in the military sphere also raises the threat of terrorism, as organized actors could easily access advanced AI-powered systems. Moreover, the fast-paced, constantly evolving nature of AI development turns military strategies into a “cat and mouse” game, where advancements are met with equally rapid countermeasures. In light of these challenges, the future of military AI must be handled with extreme caution, emphasizing robust ethical frameworks, international regulations, and stringent human oversight to prevent these technologies from destabilizing global peace.

What we need

We call for the establishment of a meaningful legally binding instrument for the use of AI-driven systems in the military requires comprehensive integration of the technical sector alongside state actors, addressing the urgent need for standardized verification protocols and trust-building mechanisms between nations. Such an instrument should define clear autonomy thresholds that specify permissible levels of independence in target selection and engagement, mandate extensive documentation of algorithmic decision processes and testing methodologies and establish explicit red lines that cannot be crossed including prohibited deployment scenarios, target categories, and operational environments. This framework should apply consistently across developing and developed nations, incorporate independent verification bodies with appropriate technical expertise to conduct regular compliance audits, and establish enforcement mechanisms with meaningful consequences for violations, all while facilitating technical data sharing and research that builds confidence between stakeholders in this domain.

These systems present an unprecedented threat to global security and human rights, and the risks they pose to non-combatants are immense. It is crucial that it

implements a robust framework of monitoring, accountability and oversight. Firstly, the states need to be bound by positive obligations to ensure the responsible use of AI in the military domain. Accountability is a fundamental aspect of this framework. We call for comprehensive mechanisms that oversee every stage of the AI system life cycle, from development and updates to transfers and research. States must ensure that any uses of AI systems are monitored, with clear reporting structures in place to address incidents promptly. Furthermore, it is vital that human operators using these systems receive thorough training and guidance to make ethical decisions in the field. The principle of meaningful human control must remain central when it comes to the use of AI in the military domain to ensure that ultimate responsibility for any actions remains with human decision makers.

Unione degli Scienziati Per Il Disarmo

[6 April 2025]

Introduction

USPID (*Unione degli Scienziati Per Il Disarmo, Union of Scientists for Disarmament*) is an association of concerned scientists – founded in 1983 and based in Italy – which promotes arms control and disarmament initiatives based on scientific understanding of risks posed by military applications of science and technology. USPID submits to the United Nations Secretary-General its views on “Artificial intelligence in the military domain, with specific focus on areas other than lethal autonomous weapons systems, and its implications for international peace and security”, in accordance with the invitation formulated in operative paragraphs 7 and 8 of Resolution [79/239](#) adopted by the UN General Assembly on 24 December 2024.

Hazards for peace and security arising from AI military applications

USPID expresses its deep concern about new hazards for peace, international security, and the respect of International Humanitarian Law (IHL) which arise on account of the ongoing and accelerating military efforts to incorporate Artificial Intelligence (AI) into multiple facets of warfare. Major sources of these hazards have been identified in current limitations of our capability to understand, predict precisely, and control the behavior of AI systems developed by machine learning methods and their interactions with other human or artificial agents. Initially identified in connection with the operation of AI-enabled Autonomous Weapons Systems (AWS), these hazards are now spreading to AI systems supporting intelligence collection, the achievement of situational awareness, and human decision-making in warfare.

Exceptionally grave concerns are raised by proposals to integrate AI in Nuclear Command, Control, and Communication (NC3) and in adjacent systems supporting nuclear decisions, and to let AI perform tasks that might directly or indirectly affect nuclear decision-making. A significant case in point is the proposal to use AI technologies in nuclear early warning and decision-support systems, which is being advanced with the expectation that AI accuracy will reduce potential errors, and its processing speed will buy more time for nuclear decision makers. However, on account of the probabilistic nature of AI information processing, one cannot exclude the risk of AI perception leading to false positives of a nuclear attack or producing perniciously unreliable recommendations given the impossibility of ensuring that the underlying models are aligned with human values and the UN overarching goal of preventing and removing threats to peace. If such mistakes occur, no matter how infrequent, large-scale and even existential implications for humanity might ensue. Accordingly, it would be imperative to proceed with time-consuming verifications of

AI responses in nuclear early warning. But these verifications would be hindered by the black-box nature of much AI information processing and by the reliance on mostly simulated data, eventually thwarting the expectation of buying more time for human decision makers.

Additional concerns are raised by proposals to exploit the rapid pace at which AI operates to speed up battlefield decision-making and targeting cycles. These proposals are fueled by the goal of gaining military advantage over potential adversaries. However, fighting at machine speed jeopardizes both the effectiveness of human oversight on AI-enabled decision support systems and the fulfilment of ethical and legal roles that are attached to human oversight of warfare action. Indeed, overly tight temporal windows for decision-making hinder effective human control over IHL threats raised by machine suggestions. Human interventions which aim at preventing inadvertent conflict escalations prompted by fighting at machine speed are similarly hampered. In addition to this, excessive speed in human-machine interactions has been identified as a factor inducing automation biases on the battlefield, and potentially skewing human decision-making even in the absence of AI failures.

Further hazards arise in connection with inherent vulnerabilities of AI learning methods and systems. Malicious manipulation of input data might be exploited to induce classification mistakes by AI systems. Moreover, poisoning attacks corrupting learning datasets may impair learning processes and the accuracy of resulting AI systems. These risks are compounded by our current inability to fully align AI systems with human goals and values, potentially causing them to deviate from strategic objectives.

Recommended actions

Mindful of these and other emerging hazards posed by the rapid adoption of AI technologies and systems in the military domain, USPID recommends

- to integrate discussion of AI in NC3 into the Non-Proliferation Treaty framework and in dedicated high-level dialogues and forums such as the Summit on Responsible Artificial Intelligence in the Military Domain (REAIM);
- to develop sustained international dialogue, good practices, and confidence-building measures concerning new and emerging risks for peace and IHL respect raised by AI warfare applications;
- to support a comprehensive and detailed inquiry aimed at identifying actual and potential AI applications in the military domain, jointly with situations of use that pose serious threats to peace, international stability, and the respect of IHL;
- to consider and investigate the need to introduce international regulations or prohibitions for those AI military applications that pose serious threats to peace, international stability, and the respect of IHL.

Women's International League for Peace and Freedom

[24 May 2024]

The Women's International League for Peace and Freedom (WILPF) has opposed war and the development of technologies of violence since its founding in 1915. WILPF has consistently condemned military spending and militarism as detrimental to human life and wellbeing. Our concerns with artificial intelligence (AI) in the military domain and its implications for international peace and security are grounded within our wider opposition to weapons, war, and violence, as well as in our opposition to patriarchal, racist, and colonial power relations that are embedded within AI technology.

While there are many perils of the military use of AI; WILPF's submission is focused on the following issues:

1. The need for human emotion, analysis, and judgement in relation to the use of force;
2. The existence of gender, racial, and other bias in AI technology and the implications for digital dehumanisation;
3. The impacts of military use of AI on privacy and personal data;
4. The environmental harms exacerbated by the military use of AI; and
5. The dangers of war profiteering and arms racing.

Due to the concerns raised in this submission and in other spaces, WILPF opposes the military use of AI. This technology, rather than placing limits on violence or harm, expands both. Governance is insufficient in the face of the profits and power the developers of these technologies seek.

In light of the concerns raised in WILPF's submission and the implications for international peace and security, WILPF urges states:

- To refrain from using AI in the military domain and to develop national laws and regulations to this end;
- To pursue a global prohibition on the military use of AI;
- To not develop autonomous weapon systems or AI-enabled weapon systems, including those that can be used to target human beings;
- To ensure protection of personal data from use by militaries, police, border enforcement, and private companies and contractors collaborating with these institutions;
- To uphold human rights and dignity online and offline; and
- To address the environmental harms generated by data centres, cloud computing, and AI by reducing the number of these centres and energy consumption and water use, which will include reducing the overall use of AI.

WILPF also urges:

- Technology companies, tech workers, scientists, engineers, academics and others involved in developing AI or robotics to pledge to never contribute to the development of AI technologies for military use;
- Financial institutions such as banks and pension funds to pledge not to invest money in the development or manufacture of AI for military use; and
- Activists, academics, affected communities, and other concerned about privacy rights, digital dehumanisation, environmental and climate justice, gender-based violence, and other issues to collaborate and strategise to oppose the development and use of AI in the military and other violent domains.

D. Scientific Community

AI, Automated Systems, and Resort-to-Force Decision Making Research Project, The Australian National University

[11 April 2025]

Introduction

This executive summary highlights policy recommendations outlined in *AI, Automated Systems, and Resort-to-Force Decision Making – Policy Recommendations: Submission to the UN Secretary General Pertaining to A/RES/79/239 (11 April 2025)*, available on the UNODA website. For a complete account of the underlying research and associated research papers, please refer to the full submission.

Underlying Research Project

This research has arisen from a **two-and-a-half-year research project (2022-2025)**, entitled *Anticipating the Future of War: AI, Automated Systems, and Resort-to-Force Decision Making*, led by Professor Toni Erskine (Australian National University) and funded by the Australian Government through a grant by the Department of Defence.

Its focus is **distinctive and critical**. While the attention of academics and policy makers has been overwhelmingly directed towards the use of AI-enabled systems in the *conduct of war* – including, prominently, on the emerging reality of ‘lethal autonomous weapons systems’ (‘LAWS’), this project has addressed the **relatively neglected prospect of employing AI-enabled tools at various stages and levels of deliberation over the resort to war**. In other words, ‘it takes us from AI on the battlefield to AI in the war-room’.¹

This research project has brought together **leading scholars and practitioners** working on different aspects of international politics and security, strategic and defence studies, and artificial intelligence (AI) to contribute to a multi-disciplinary study and set of **policy recommendations on the risks and opportunities of introducing AI, machine learning (ML), and automated systems** into state-level decision making on the **initiation of war**. Our interventions are made from the perspectives of political science, international relations, law, computer science, philosophy, sociology, psychology, engineering, and mathematics.

Project participants presented and discussed their research at two workshops (June 2023 and July 2024) at the Australian National University (ANU), convened by Professor Toni Erskine and Professor Steven E. Miller (Harvard). Participants also received feedback on their initial research-based policy recommendations from senior Australian Government delegates from the federal civil service as part of a one-day policy roundtable (July 2024) at the ANU.

‘Four Complications’

For all the potential **benefits** of AI-driven systems – which are able to analyse vast quantities of data, make recommendations and predictions by uncovering patterns in data that human decision makers cannot perceive, and respond to potential attacks with a speed and efficiency that we could not hope to match – challenges

¹ T. Erskine and S. E. Miller, ‘AI and the Decision to Go to War: Future Risks and Opportunities’, *Australian Journal of International Affairs*, Vol. 78: 2 (2024), pp. 135–147 (p. 138).

abound. Through this project, we have sought to address **four thematic ‘complications’** that we propose will accompany the gradual infiltration of AI-enabled systems in **decisions to wage war**:²

- **Complication 1** relates to the displacement of human judgement in AI-driven resort-to-force decision making and possible implications for deterrence theory and the unintended escalation of conflict.
- **Complication 2** highlights detrimental consequences of automation bias, or the tendency to accept without question computer-generated outputs – a tendency that can make human decision makers less likely to use (and maintain) their own expertise and judgement.
- **Complication 3** confronts algorithmic opacity and its potential effects on the democratic and international legitimacy of resort-to-force decisions.
- **Complication 4** addresses the likelihood of AI-enabled systems impacting organisational structures and chains of command, whether degrading or enhancing strategic and operational decision-making processes.

Contributors to this project have explored these proposed complications in the context of either **automated self-defence** or the use of **AI-driven decision-support systems (DDS)** that would inform human resort-to-force deliberations. We have identified risks and opportunities of using AI-enabled systems in these contexts and make recommendations on how risks can be mitigated and opportunities promoted.

Complication 1: Displacement of human judgement

AI in Nuclear Crisis Decision Making

One key area of research undertaken in response to this complication is the nuanced interplay between AI and human decision making in the high-stakes context of **nuclear crisis management**. Risks (including the increased fragility of nuclear deterrence relationships, crisis signalling becoming more complex, and unintended escalation) have been explored in two broad areas: i) automation in military deployments, or taking the human ‘out of the loop’ in the decision to use nuclear or strategic non-nuclear weapons (SNNW); and, ii) the integration of AI into human decision-making (particularly in early warning threat assessments). Although much of this research has focused on risks, **novel benefits** of introducing AI-driven decision-support systems (DSS) into human-led nuclear crisis management have also been proposed.

Policy Recommendations:

- **Always incorporate human-in-the-loop safeguards:** Ensure AI systems in nuclear command and control are always overseen by human operators and that human decision-making remains central to determining when and how nuclear-weapon states resort to the use of their arsenals.
- **Promote a holistic approach to AI-safety:** AI safety should account for both technical and socio-technical dimensions. Assess safety challenges in AI-enabled DSS comprehensively, including issues of security, trust, and liability.
- **Broaden the scope of risk assessments:** Apply risk assessments relating to the deployment of AI and ML not only to obvious areas such as nuclear launch orders, but also to less obvious areas such as early warning intelligence

² For an account of these ‘four complications’, see T. Erskine and S. E. Miller, ‘AI and the Decision to Go to War: Future Risks and Opportunities’, *Australian Journal of International Affairs*, Vol. 78: 2 (2024), pp. 135–147 (pp. 139–40).

assessments (including by non-nuclear allies) and SNNW capabilities (including by non-nuclear allies).

- **Restrict the use of AI-assisted warning data:** The key to balancing the benefits of incorporating AI into early warning against the risks is limiting what AI-assisted warning data is used for. In AI research, prioritise tasks such as calculating effective evasive manoeuvres in the event of an attack and using pattern recognition and anomaly detection to improve arms control verification.
- **Pursue informal arms control and confidence-building:** Advance informal measures such as regular dialogue, red-line agreements, and information-sharing mechanisms. Expand unilateral initiatives like moratoriums where feasible.
- **Explore AI's potential to promote empathy and enhance decision making:** Decision makers must exercise ‘security dilemma sensibility’ (SDS) in times of crisis. Decision makers and diplomats exercise SDS when they are open to the possibility that the other side is behaving the way they are because they are fearful and insecure, and crucially, recognize the role that their own actions may have played in this. Explore ways that the balanced integration of AI and human judgement could enhance SDS during nuclear crises by promoting empathy and trust.

AI Mistakes in the Resort to Force

Another area addressed in relation to this complication is **state responsibility** when **errors** occur in AI-driven or autonomous systems involved in resort-to-force decisions. Such errors may arise from poor system training, data poisoning by adversaries, or two AI-driven systems interacting in unintended ways. It is essential to develop legal standards and practices that reduce the risk of unintended conflict resulting from such failures.

Policy Recommendations:

- **Adopt robust security and cyber hygiene:** States should adopt robust protections against AI data poisoning and cyber attacks to meet jus ad bellum standards of good faith and reasonable conduct.
- **Clarify legal guidelines on delegating the use of force to autonomous systems:** Senior leadership within states should set clear domestic legal standards regarding when and how autonomous systems may be authorised to use force.
- **Commit to transparency in after-action reviews:** States should commit to being transparent and deliberate about after-action reviews of any AI errors that occur in the field, potentially drawing on civilian casualty review processes as a model.

Complication 2: Automation bias

Our research in response to the second complication focuses on the relationship between human actors and **AI-driven DSS** in resort-to-force decision making. It includes a detailed survey-based study of **military trust in AI** during strategic-level deliberations and a robust account of the importance of ensuring that there are human ‘**experts-in-the-loop**’ when AI-driven systems contribute to decisions on war initiation. This body of work also addresses the **benefits** of employing DSS to **enhance our cognitive capacities** in strategic decision making and, conversely, uncovers the potential **dangers** of such reliance if these systems **dull our sensitivity**.

to the tragic qualities of war or contribute to the **erosion of restraint** by creating the illusion that they replace us as responsible actors.

Policy Recommendations:

- **Consider the multidimensionality of trust:** Recognize that soldiers' trust in AI is not a forgone conclusion. Rather, it is complex and multidimensional, and further complicated by biases, uncertainty, and lack of education.
- **Interrogate norm compliance:** In terms of governance, explain how policies on increasingly autonomous capabilities coincide or diverge from international norms and laws informing their use.
- **Embed experts in decision structures:** Enshrine an 'expert-in-the-loop' organisational structure – i.e., high-level experts as core decision makers.
- **Prohibit automation:** Prohibit automation of resort-to-force decisions.
- **Increase AI literacy of domain experts:** Provide and require basic technical training for high-level domain experts so they understand the logics of AI and can thus incorporate AI decision inputs from an informed position.
- **Provide on-going, substantive training for domain experts:** Sustain substantive training for, and assessment of, high-level experts to bolster and ensure substantive competencies.
- **Regulate non-autonomous AI:** While autonomous AI agents, e.g., lethal autonomous weapons systems (LAWS), need regulation, so do non-autonomous AI systems, which leave humans vulnerable to new forms of influence, moral and cognitive atrophy, and undermined responsibility.
- **Design AI-driven DSS to promote more accurate perceptions of their capacities:** Ensure AI-driven DSS are not easily mistaken for responsible agents in themselves by avoiding anthropomorphism, building in warnings about system limitations, and incorporating features that emphasise human agency and accountability.

Complication 3: Algorithmic opacity

Our research in response to the third complication addresses **how the lack of transparency of AI-driven decision making can threaten the legitimacy** of AI-informed decisions on the resort to force. This body of work includes original research on **large language models (LLMs)** and their potential to exacerbate existing **pathologies in intelligence analyses**. It also examines the role that the '**architecture of AI**' and its hidden vulnerabilities play in deliberations surrounding the resort to force. Moreover, research within this pillar conceives of military decision-making institutions as '**complex adaptive systems**' – a conceptual framework that yields a range of insights, including that human-machine teams possess a form of '**cognitive diversity**' that could be leveraged for **more efficient decision-making**, but also exploited to poison information flows, and that technical explanations for algorithmic opacity will not solve accountability concerns.

Policy recommendations:

- **Develop policy to limit epistemic pathologies of LLMs:** States should clearly determine defence and intelligence policy towards either a) procurement of LLMs, b) state development of LLMs, or c) a combination of both. They should use this guidance to develop policy which seeks to limit the epistemic pathologies of LLMs in autonomous decision-making.

- **Commit to sector-wide procurement guidelines and oversight of generative AI tools** used in decision-making chains.
- **Commit to regulating data markets** and access to those markets through alliance relationships.
- **Promote understanding of the tech ecosystem and its fragilities:** Increase understanding of the inherent interdependencies and vulnerabilities in the tech ecosystem, including by creating technology literacy training programs designed specifically for politicians and policy, intelligence, and military leaders.
- **Invest in research** to develop a comprehensive picture of the architecture – physical and digital – that underpins AI, including critical dependencies and vulnerabilities and how access and power are distributed.
- **Invest in research on social media** and its impact on functions of government, including its potential to disrupt democracies, facilitate foreign interference, and influence decision making on the use of force.
- **Recognize AI's current influence:** Significantly increase awareness of government reliance on the architecture of AI, especially for critical government functions, including resort-to-force decision making.
- **Invest in research and development** to maximize the benefits of human-machine cognitive diversity.
- **Implement responsible AI governance programs** that carefully balance accountability with operational efficiency.
- **Perform regular red-team exercises** to ensure that the integration of AI in decision-making institutions does not induce systemic blind spots and vulnerabilities in military decision-making.

Complication 4: Impact on organisational structures

Our research regarding the fourth complication explores both the **beneficial and damaging effects that AI-driven systems can have on institutional structures** in the context of resort-to-force decision making. Studies focus on how AI-driven DSS **can improve ‘adaptive culture’** within military organisations, thereby improving wartime decisions, and how the urgent need to **upgrade AI literacy and educate human analysts** should lead us to **reform institutional structures and cultures**. The novel notion of **‘proxy responsibility’** is proposed as an institutional response to ensure that responsibility can be meaningfully assigned to humans for resort-to-force decisions that are informed by AI systems. Moreover, original research highlights the significance of the **neglected category of AI ‘integrators’** – sandwiched between the ‘developers’ and ‘users’ of AI within organisational structures – when it comes to strategic military applications of AI.

Policy recommendations

- **Set (and evolve) measures of effectiveness.** If AI-enabled adaptive capacity is to work effectively, measures of military effectiveness must guide which direction adaptation might take. Establish such measures at the tactical (battlefield) and strategic (war-room) levels to guide development and implementation of AI-enabled adaptation.
- **Know where adaptation relevant data is found, stored and shared.** An enhanced adaptive stance in military institutions must have enhanced data

awareness as a foundation. Data awareness and management must become one of the basic disciplines taught to military personnel.

- **Scale AI support from individual to institution.** There is unlikely to be a one-size-fits-all algorithm or process that can enhance learning and adaptation at every level of military endeavours. Create a virtual ‘arms room’ of adaptation support algorithms as part of an institution-wide approach to adaptation.
- **Routinely question AI-enabled outputs:** Build mindsets, protocols, institutional cultures, and inter-agency structures in ‘normal’ pre-crisis times to routinely question AI-enabled output from human-machine teams.
- **Institute an advisory body:** In order to support the notion of ‘proxy responsibility’ as an institutional response to ‘responsibility gaps’ when decisions on war initiation are informed by AI-enabled systems, establish and/or strengthen state-level ‘AI departments’. These departments would integrate technical, political, and ethical competence and expertise and advise on resort-to-force decision-making processes.
- **Support research on AI integration:** Fund research on the integration of AI in strategic decision-making.
- **Provide standards:** Outline minimum standards for the responsibilities of AI developers and integrators.
- **Facilitate inter-group discussions** between developers, integrators and users during development, integration, and longer-term maintenance processes.
- **Create accountability guidelines:** Provide well-defined guidelines and rules indicating who is accountable if something goes ‘wrong’.

Queen Mary University of London, T.M.C Asser Institute, University of Southern Denmark, University of Utrecht

[11 April 2025]

Views of members of the scientific community and civil society; specifically, we are a group of academics with expertise in ethical, legal and political dimensions of military Artificial Intelligence and herewith put forward our shared views pursuant to resolution [79/239](#) “Artificial intelligence in the military domain and its implications for international peace and security” adopted by the General Assembly on 24 December 2024, in accordance with the request of the UN Secretary-General contained in Note Verbale ODA/2025-00029/AIMD.

Introduction:

The rapid advancement and integration of AI technologies into targeting operations have sparked ongoing debates surrounding their ethical, legal, and operational implications. Over the past decade, the discourse on AI in warfare has largely centered on autonomous weapon systems (AWS),¹ driven in part by the initiation of discussions in 2013 and the formalization of a regulatory process under the UN Convention on Certain Conventional Weapons (CCW) and the Group of Governmental Experts on Lethal Autonomous Weapons Systems (GGE LAWS),

¹ The latest definition of AWS from the CCW GGE LAWS Rolling Text (26 November 2024): “A lethal autonomous weapon system can be characterized as an integrated combination of one or more weapons and technological components that enable the system to identify and/or select, and engage a target, without intervention by a human user in the execution of these tasks.” On file with authors.

which exclusively focuses on lethal AWS.² However, the increasing integration of AI-based decision-support systems (AI-DSS) into targeting practices³ introduces new layers of complexity that demand closer attention from a broad range of stakeholders. This submission responds to that need, structured around three key components: (1) a brief overview of how AI-DSS are currently used in targeting decisions; (2) an analysis of key concerns, including how these systems shape the potential exercise of human judgement and control and underline fundamental gaps in global governance; and (3) a concluding set of recommendations.

1. Overview of AI-DSS and the joint targeting cycle

Defined as “the process of selecting and prioritizing targets and matching the appropriate response to them, considering operational requirements and capabilities,”⁴ targeting is a core military function at the very heart of warfare. While the potential range of use cases for AI-DSS in military decision-making is broad, in targeting, AI-DSS can be understood to serve as **tools** that use AI techniques to collect and analyze data, provide information about the operational environment as well as actionable recommendations, with the aim of aiding military decision makers in evaluating factors relevant to legal compliance such as taking precautions and ensuring proportionality in attacks.

More specifically, AI-DSS are increasingly integrated across multiple phases of the joint targeting cycle (JTC), including within target development and prioritization, capabilities analysis, and mission execution. The JTC is a reflective example of a structured process used by military forces to identify, evaluate, and engage targets while ensuring compliance with operational, legal, and ethical standards,⁵ generally consisting of six (non-linear) phases:

1. **End-State and Commander’s Objectives:** Defining strategic military goals and desired outcomes.

² For a brief overview of some of the latest developments of the GGE LAWS see Jeroen van den Boogaard, *Warning! Obstacles Ahead! The Regulation of Autonomous Weapons Systems in the GGE LAWS*, *Opinio Juris*, 4 March 2024 found at: <https://opiniojuris.org/2024/03/04/warning-obstacles-ahead-the-regulation-of-autonomous-weapons-systems-in-the-gge-laws/>.

³ There have been several reported uses of AI-DSS by Israel in Gaza and potentially in Lebanon, by both Ukraine and Russia in the ongoing conflict, and by the United States in its actions against Houthi rebels in the Red Sea and in Yemen, to name a few. For a comprehensive overview of literature in this space, see e.g., Anna Nadibaidze, Ingvild Bode, and Qiaochu Zhang, *AI in Military Decision Support Systems, A Review of Developments and Debates*, Centre for War Studies, University of Southern Denmark, November 2024. Found here: <https://www.autonorms.eu/ai-in-military-decision-support-systems-a-review-of-developments-and-debates/>.

⁴ United States Department of Defense, *Dictionary of Military and Associated Terms*, March 2017, found at: <https://www.tradoc.army.mil/wp-content/uploads/2020/10/AD1029823-DOD-Dictionary-of-Military-and-Associated-Terms-2017.pdf>.

⁵ Michael Schmitt et al, *Joint and Combined Targeting: Structure and Process*, Chapter 13 in Jens David Ohlin (ed) *Weighing Lives in War* (Oxford, 2017). See also, Jessica Dorsey and Marta Bo, *AI-Enabled Decision-Support Systems in the Joint Targeting Cycle: Legal Challenges, Risks, and the Human(e) Dimension*, forthcoming 2025, *International Law Studies*. “Targeting generally involves four key steps: (1) objectives and guidance, (2) planning, (3) execution, and (4) assessment. Encapsulating these four key steps, the United States and NATO outline their targeting processes through similar six-phase cycles [addressed in this submission]. As the reader can discern, different states employ different doctrines for targeting. What is important ... is not necessarily the specific labels for various steps followed by any given state, but rather how and when compliance with the principle[s of IHL are] incorporated into the targeting process.”

2. **Target Development and Prioritization:** Identifying, verifying/validating, and prioritizing targets based on intelligence and mission goals.
3. **Capabilities Analysis:** Assessing the available strike options and their effectiveness.
4. **Force Assignment:** Allocating specific military assets (e.g., airstrikes, artillery, cyber operations) to engage the target.
5. **Mission Execution:** Carrying out the targeting operation while ensuring compliance with relevant laws and the rules of engagement.
6. **Assessment:** Evaluating the effectiveness of the operation and adjusting for future operations, if necessary.

Within this framework, AI DSS are assumed to serve primarily as informational and analytical tools which support human decision-making rather than supplant it. However, this assumption and framing obscures how AI-DSS influence human cognitive processes within the JTC. This impact on human decision-making is often underestimated and remains insufficiently examined, leaving critical discussions about the role of AI-DSS largely absent from current policy debates.

2. Analysis of Key Concerns

(a) (Meaningful) Human Judgement and Control

AI-DSS are often portrayed as enhancing human decision-making and the quality of decisions therein. The perception of AI-DSS as mere subsidiary tools has led to a narrative that the integration of AI-DSS poses fewer challenges than AWS, given that these systems do not directly “engage” targets (i.e., they do not have an inherent capability to directly carry out the use of force) and are tools that assist human commanders. The outputs are ostensibly ultimately reviewed through (several layers of) human oversight, such as processes of verifying and validating targets using additional intelligence sources.⁶ As a result, errors or inaccuracies in AI-DSS outputs are often seen as non-critical, based on the assumption that robust human oversight and appropriate control will compensate for them. However, closer examination reveals that this control is frequently superficial, offering only the appearance of, rather than actual meaningful, or context-appropriate, human judgement and control.

This is because AI-DSS structure and condition the quality of human control and oversight and limit the ways control and oversight can be exercised. The use of AI-DSS creates a shared decision-making space between human military personnel and AI technologies. States appear to have recognized and focused on many of the advantages of this shared decision-making space for military personnel, i.e., how the use of AI-DSS advances human decision-making through offering data-driven insights. But using AI-DSS also delimits the capacity to exercise human oversight and control because of the technologies’ complexity and the increased speed (and therefore scale) it can bring to decision-making processes. Rather than supporting human oversight, using AI-DSS may risk humans becoming little more than reactive cogs in socio-technical systems.⁷ Moreover, this configuration risks amplifying adverse human biases, such as automation bias, anchoring bias, or cognitive action bias, to the detriment of exercising qualitatively high levels of human control.⁸

⁶ Alexander Blanchard and Laura Bruun, *Automating Military Targeting: A Comparison Between Autonomous Weapon Systems and AI-Enabled Decision Support Systems*, Stockholm International Peace Research Institution (SIPRI) forthcoming 2025 (draft on file with authors).

⁷ Ingvild Bode, *Human-Machine Interaction and Human Agency in the Military Domain*, Policy Brief No. 193 (Waterloo, ON: Centre for International Governance Innovation, 2025).

⁸ Dorsey and Bo, *supra* n. 5.

Considering AI-DSS as a distinct form of technology therefore reveals significant challenges associated with military AI and human oversight, challenges that extend beyond those that arise when simply integrating the technology in weapon systems.

Recent conflicts have shown the risks associated with AI-DSS being employed in critical functions, such as target selection and even nomination, and their conditioning and constraining of human involvement, affecting the fulfilment of core legal obligations embedded within the JTC. The use of AI-DSS raises fundamental concerns about whether human decision makers can retain adequate cognitive autonomy over the JTC process or whether humans will become overly reliant on algorithmic outputs for critical judgements in the context of armed conflict.⁹ Consequently, there are significant legal concerns regarding the effects of such systems on decision-making processes and use of force decisions and ability for users to comply with IHL obligations, especially with respect to the obligation to take all feasible precautions to minimize civilian harm to the greatest extent possible in attack and comply with the principles of distinction and proportionality.¹⁰

Importantly, these concerns are not new. There is extensive debate around how to preserve meaningful human judgment and human agency when conducting IHL-evaluative legal assessments, in the context of AWS. These discussions – which include expert analysis on accountability, human-machine interaction, automation bias, and the effect of AI systems on legal and ethical reasoning¹¹ – provide valuable lessons that must inform discussions around military AI and specifically the use of AI-DSS in military contexts.

(b) AI-DSS: Understudied, Under-Addressed and Unregulated

Framing AI-DSS as mere tools, has led to an underestimation and lack of analysis on the way their use affects the cognitive decision-making process within the JTC. The relative lack of attention paid to AI-DSS so far can partly be attributed to the fact that such systems are seen to be used with a human *in or on* the loop framework, with their outputs ostensibly reviewed by one or more individuals during the targeting process. As a result, current understandings of AI-DSS use appear to align with widely supported principles of human control and oversight. However, this gap in the debate

⁹ *Ibid*; see also Anna Nadibaidze, Ingvild Bode, and Qiaochu Zhang, *AI in Military Decision Support Systems, A Review of Developments and Debates*, Centre for War Studies, University of Southern Denmark, November 2024. Found at: <https://www.autonorms.eu/ai-in-military-decision-support-systems-a-review-of-developments-and-debates/>.

¹⁰ Article 57 of the First Additional Protocol to the Geneva Conventions. See also Dorsey, Bo supra n. 5 (on AI-DSS and their effects on the principle of precautions); Jessica Dorsey, *Proportionality under Pressure: The Effects of AI-Enabled Decision Support Systems, the Reasonable Commander Standard and Human(e) Judgment in Targeting*, forthcoming *International Review of the Red Cross* (2025) (on AI-DSS and their effects in the context of IHL proportionality assessments).

¹¹ Marta Bo, *Autonomous Weapons and the Responsibility Gap in light of the Mens Rea of the War Crime of Attacking Civilians in the ICC Statute*, 19 *Journal of International Criminal Justice* 2021; Bo, M., Bruun, L. and Boulanin, V., *Retaining Human Responsibility in the Development and Use of Autonomous Weapon Systems: On Accountability for Violations of International Humanitarian Law Involving AWS* (SIPRI: Stockholm, Oct. 2022), p. 41; Boulanin, V., Bruun, L. and Goussac, N., *Autonomous Weapon Systems and International Humanitarian Law: Identifying Limits and the Required Type and Degree of Human-Machine Interaction* (SIPRI: Stockholm, June 2021), p. 54; and Bruun, L., Bo, M. and Goussac, N., *Compliance with International Humanitarian Law in the Development and Use of Autonomous Weapon Systems: What Does IHL Permit, Prohibit and Require?* (SIPRI: Stockholm, Mar. 2023), p. 24. Elke Schwarz, “The (im)possibility of meaningful human control for lethal autonomous weapons systems,” *Humanitarian Law and Policy*, 29 August 2018, found at: <https://blogs.icrc.org/law-and-policy/2018/08/29/im-possibility-meaningful-human-control-lethal-autonomous-weapon-systems/>.

is also caused by a lack of transparency around how specific AI-DSS function, and a consistent failure to comprehensively examine how they are being used in practice.

Additionally, the persistent focus on AWS at the expense of AI-DSS obscures the growing reliance on AI in shaping operational and strategic outcomes. Unlike AWS, which have been debated in the framework of the CCW for the past decade, AI-DSS lack a comparable institutional platform. Attention to AI-DSS remains scattered across various initiatives but these efforts have yet to provide the dedicated regulatory focus or coordination needed.

3. Recommendations:

- i. **Reassert** the central role of human cognitive and legal reasoning in military operations by implementing safeguards that ensure key legal assessments remain grounded in human(e) judgment. Leverage existing insights from debates on AWS and research on human-machine teaming and human-computer interaction to inform discussions on AI-DSS.
- ii. **Recognize** and address the incremental effects of AI-DSS design and use on human cognitive reasoning and critical deliberation. Promote awareness and attentiveness as a crucial part of reasserting and strengthening the exercise of human agency in targeting decision-making.
- iii. **Reinforce** calls for greater attention to the implications of AI-DSS in armed conflict. Utilize platforms such as the UN General Assembly's First Committee on Disarmament and International Security and the Global Commission on the Responsible Use of AI in the Military Domain to foster inclusive and complementary discussions on the associated risks and systemic changes AI-DSS introduce.

United Nations Institute for Disarmament Research

[11 April 2025]

Artificial intelligence (AI) is rapidly transforming the military domain and profoundly influencing international peace and security. Initiatives such as the summits on Responsible AI in the Military Domain (REAIM) and the Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, while not being universal processes, have significantly elevated international attention on the military applications of AI. In particular, they have moved the debate beyond lethal autonomous weapon systems (LAWS) and have successfully highlighted the multifaceted impacts of AI, fostering broader international policy engagement. Building on the political momentum generated by these initiatives, resolution [79/239](#) adopted by the United Nations General Assembly in December 2024 represented a significant milestone as the first UN resolution on AI in the military context and has offered Member States, international and regional organizations and the multi-stakeholder community the opportunity to share their views on opportunities and risks.

For many years, the United Nations Institute for Disarmament Research (UNIDIR) has played an important role in shaping and informing discussions on the broader impact of AI in the military domain, both within and beyond applications of this technology in weapon systems. It has undertaken research, facilitated multilateral dialogues, and offered policy insights that underline AI's transformative potential for international peace and security. This policy note draws from all the work conducted to summarise opportunities and risks and to offer a potential roadmap for future policy action.

The international community can now shape how AI is used in the military domain, putting principles of responsible AI at the core. A central challenge is the

complexity of defining the “military domain”. States and regions interpret the scope of this domain differently based on their unique security landscapes, realities and operational practices. For some countries, military roles extend to internal security tasks such as policing, border control, combating organized crime, protection of critical infrastructure or humanitarian relief in response to natural disasters. Others maintain a stricter definition, limiting military functions to battlefield engagements. These variations, rather than serving as barriers, offer important context for multilateral discussions. International governance frameworks should remain flexible and inclusive, acknowledging and adapting to diverse national and regional security perspectives.

In the many operational contexts within the military domain, AI acts as a force multiplier across several military tasks, including command and control (C2), information management and intelligence, advanced autonomy, logistics, training and simulation, and organizational and support functions. In C2, AI enhances the speed and quality of decision-making, thereby helping commanders rapidly analyse battlefield scenarios. It has the potential to improve adherence to international humanitarian law (IHL), for example by integrating detailed proportionality and other legal assessments. AI-driven intelligence tools analyse large volumes of data at speed, and so improve situational awareness and threat detection. In logistics, AI optimizes supply chains and predictive maintenance, enhancing operational readiness and improving the sustainability of military operations over time. AI further supports advanced autonomy in drones, cybersecurity, and operations in the information domain. Training and simulation benefit from AI by creating personalized, realistic synthetic environments and scenarios. In short, if developed, deployed and used responsibly, AI could increase operational effectiveness while offering new ways to mitigate risks and reduce harm.

However, integrating AI in military contexts also presents significant risks and challenges – technological, security, legal, policy and ethical.

Technologically, military AI systems face issues related to the quality, availability and inherent biases of data. These may lead to unpredictable and potentially harmful outcomes, including violations of international law. The “black box” nature of AI systems, often coupled with their adaptiveness and highly context-dependent nature, complicates trustworthiness assessments and may, at times, challenge the conduct of effective investigations into alleged violations of IHL. Cybersecurity vulnerabilities also expose AI systems to adversarial attacks, requiring stringent security measures.

Security challenges include risks of miscalculation and unintended escalation, particularly through AI-enabled rapid decision-making processes and AI-enabled autonomy, which may result in escalatory responses. The potential for an AI arms race exacerbates international and regional tensions, possibly leading to destabilizing outcomes similar to historical arms competitions. The proliferation of AI technologies to non-state actors further complicates threat landscapes and necessitates robust life cycle management of military AI systems. Additionally, AI-generated disinformation threatens societal stability by undermining trust in information and can have a direct impact on military operations.

Legal challenges revolve around ensuring compliance with international law, particularly IHL and international human rights law. Key debates focus on, among other things, accountability and both state and individual responsibility for AI-driven actions, especially regarding lethal decisions. States diverge on whether existing legal frameworks are sufficient or if new, specialized regulations are required. Beyond international law, ethical considerations emphasize maintaining human judgment in critical decision-making and preventing societal biases from infiltrating AI systems. The latter requirement calls for greater diversity and inclusivity in AI development. Additionally, bridging gaps between government, academia and the private sector remains challenging yet crucial for effective governance.

Addressing these challenges requires a comprehensive road map with actions at the multilateral, regional and national levels.

Multilaterally, establishing a United Nations-led comprehensive platform that enables a regular institutional dialogue to address military AI's broader implications on international peace and security is key as it would provide an institutional framework to advance policy discussions. This platform could build on the existing internationally developed AI principles and frameworks, such as UNESCO's recommendations or the commitments made in the Global Digital Compact (e.g. safe, secure and trustworthy AI) and further refine them for application in the military domain. These principles could be further developed into voluntary norms of responsible behaviour in the development, deployment and use of AI in the military domain and provide a solid foundation for future multilateral instruments. In addition, such platform could be leveraged to develop practical confidence-building measures (CBMs), lead inclusive multi-stakeholder engagement, and deliver global capacity-building programmes that enhance global security via transparency, cooperation and predictability.

Regionally, existing organizational frameworks can be used to tailor CBMs and guidelines that reflect local security contexts. Cross-regional dialogues would facilitate mutual learning, prevent information silos, and include diverse perspective which would encourage globally coherent responses.

Nationally, states should develop comprehensive AI strategies that detail vision, priorities and governance frameworks, ensuring compliance with international norms and ethical standards. Robust governance structures (e.g., dedicated AI steering committees and ethics boards), alongside iterative legal reviews, would enhance accountability and safety. Transparent communication and clearly defined accountability protocols would further support responsible AI implementation. High standards of data governance, life cycle management approaches, rigorous training programmes and updated military operational guidelines complete these proposed national measures, ensuring the responsible integration of AI in the military domain.

Table A below provides an overview of the proposed roadmap for policy action.

Table A: A roadmap for future policy action

<i>Level</i>	<i>Action</i>	<i>Rationale</i>
Multilateral	<p>Establish a multilateral process under United Nations auspices to provide a comprehensive platform for discussion on military applications of AI and their impact on international peace and security. This process could be leveraged to:</p> <ul style="list-style-type: none"> a. Develop a set of overarching, core principles of responsible AI in the military domain to help align national efforts and reduce risk. b. In the future, further develop these core principles into international voluntary norms or guidelines for 	<p>Collectively, these multilateral actions aim to foster cooperation, set common rules and share knowledge on military AI at the international level with a view to increasing predictability.</p> <p>They aim to shape the global landscape so that all states move towards safer and more transparent integration of AI in the military domain, thereby reducing the risks.</p> <p>While clustered under a single umbrella recommendation, each of the actions above could be implemented on its own, although their mutually</p>

<i>Level</i>	<i>Action</i>	<i>Rationale</i>
	<p>responsible state behaviour in the development, deployment and use of AI in the military domain. These guidelines could take the form of a code of conduct or a political declaration supplemented by more technical instruments as required (e.g., on AI assurances, and robust protocols for testing and evaluation).</p> <ul style="list-style-type: none"> c. Develop confidence-building measures (CBMs) for military AI. States could agree on and implement practical CBMs to increase transparency and trust regarding AI in the military domain. d. Promote multi-stakeholder engagement in support of multilateral policy action. e. Develop and implement a coherent capacity-building programme. 	reinforcing nature would amplify the impact achieved if they are implemented in combination.
Regional	<p>Leverage regional and subregional organizations and dialogues to discuss the issue of AI in the military domain.</p> <p>Regional and sub-regional organizations could:</p> <ul style="list-style-type: none"> a. Develop region-specific CBMs, norms or guidelines that reflect local contexts. b. Set up networks for information-sharing on AI-related best practices suited to their security landscape. c. Develop joint AI-development projects, aligning operational, legal and technical requirements. <p>Initiate cross-regional dialogues</p> <p>Initiate cross-regional dialogues on AI, where two or more regional groups exchange lessons and possibly align their approaches.</p>	<p>Regional and subregional approaches allow tailoring to specific security realities and threat perceptions, which could lead to concrete results that are more aligned with specific needs.</p> <p>In addition, regional and subregional approaches could be leveraged to inform and shape global dialogues and strengthen context-specific capacity-building.</p> <p>Cross-regional dialogue can be a useful tool to enable mutual learning and avoid echo chambers.</p>

<i>Level</i>	<i>Action</i>	<i>Rationale</i>
National	<p>Implement a comprehensive approach to AI governance in the military domain to include the following actions:</p> <ul style="list-style-type: none"> a. Develop a comprehensive national strategy or policy on AI in security and defence. b. Establish robust governance structures and review processes. c. Implement transparency and accountability measures d. Implement robust data practices and governance frameworks for all military AI applications. e. Manage AI capabilities throughout their entire life cycle – from design and development, through testing and deployment, to updates and decommissioning – with continuous risk assessments and mitigation at each stage. f. Invest in human capital and training by developing extensive training programmes for military personnel on AI and cultivating a new generation of AI-literate officers and specialists. This includes not only technical training but also training on the ethical and legal aspects of AI use in operations. g. Review military operational guidelines to strengthen AI governance in military contexts, including military documentation (e.g. doctrines, standard operating procedures and others), and rules of engagement. 	<p>A national strategy clarifies roles and responsibilities, and provides a clear direction for the development, acquisition, integration and use of AI in the military domain.</p> <p>Dedicated structures provide focus and accountability. They create effective checkpoints that AI projects must pass and comply with consistently (e.g., ethical approval, legal clearance, safety certification), reducing chances of unsafe or unlawful deployment.</p> <p>Transparency builds public trust and international confidence that a state is using AI responsibly.</p> <p>Accountability ensures that the presence of AI does not create a vacuum of responsibility – maintaining the ethical and legal norm that humans are accountable for military actions.</p> <p>By prioritizing robust data governance and the provision of the necessary infrastructure to enable it, militaries can improve the performance and trustworthiness of their AI systems and reduce error rates.</p> <p>A life cycle view ensures that safety and compliance are ongoing commitments reducing chances of failure in the field and ensuring that accountability is maintained throughout the system's use.</p> <p>Human expertise and judgment remain critical. Training reduces misuse and enables more effective human-machine teaming.</p> <p>Existing military governance tools and instruments can be used to strengthen the governance of AI in the military domain at a more practical, tactical level, thereby offering an</p>

<i>Level</i>	<i>Action</i>	<i>Rationale</i>
		impactful complement to the highest levels of governance and the associated obligations emanating from international, regional and national laws and regulations.

E. Industry

Microsoft

[24 May 2024]

Microsoft welcomes the opportunity provided by the United Nations General Assembly resolution [A/RES/79/239](#) on “Artificial Intelligence in the Military Domain and its Implications for International Peace and Security”, and UNODA’s invitation to share perspectives on the opportunities and challenges posed to international peace and security by the application of artificial intelligence (AI) in the military domain, with specific focus on areas other than lethal autonomous weapons systems.

Our perspectives reflect Microsoft’s deep commitment to our Responsible AI Principles and our Secure Future Initiative, emphasizing cybersecurity, safeguarding international norms, and promoting trust in technology, and our active participation in multi-stakeholder initiatives including the UNIDIR-led Roundtable for AI, Security, and Ethics (RAISE).

I. Opportunities

Microsoft recognizes substantial opportunities in responsibly applied AI within the military domain, particularly:

- *Enhancing cybersecurity and defense capabilities:* AI significantly strengthens cybersecurity defenses by automating threat detection, enabling faster and more accurate responses to cyber threats. Technologies such as Microsoft Security Copilot illustrate the transformative potential of AI in defense, empowering cybersecurity professionals to identify and mitigate risks efficiently. Initiatives like Microsoft’s Zero Day Quest and collaboration with MITRE ATT&CK demonstrate proactive industry efforts to enhance global cybersecurity preparedness and resilience.
- *Broad spectrum of military applications:* Beyond cybersecurity, responsibly designed AI can significantly enhance efficiency and effectiveness across logistics, command and control systems, intelligence processing, military training, peacekeeping, humanitarian assistance, and disaster relief operations. Diverse applications underscore AI’s transformative potential beyond combat scenarios alone.
- *Improving compliance with international humanitarian law:* AI technologies should improve the accuracy and effectiveness of targeting processes, aiding militaries to better adhere to principles of distinction, proportionality, and necessity. AI should significantly enhance protections for civilians and civilian infrastructure, thereby reducing unintended collateral damage in conflict.
- *Capacity building and international cooperation:* The adoption of AI in the military domain presents opportunities for global knowledge-sharing and

capacity-building initiatives. International partnerships should support developing nations by sharing security capabilities, knowledge, and best practices, thus bridging technological divides and fostering global stability.

II. Challenges

Microsoft equally acknowledges significant challenges and risks associated with AI applications in the military domain:

- *AI-enhanced cyber threats:* AI has escalated cyber threat capabilities, empowering state-sponsored and criminal actors to carry out increasingly sophisticated cyber operations. These AI-driven threats include advanced phishing campaigns, automated exploitation of vulnerabilities, and adaptive malware, significantly increasing global cybersecurity risks.
- *Risks of escalation and miscalculation:* Integrating AI into military decision-making risks unintended escalation and/or miscalculation. Rapid, automated decision-making processes may inadvertently lower conflict thresholds, amplifying risks of destabilization or accidental conflict.
- *Proliferation and uncontrolled diffusion:* Uncontrolled diffusion, especially through open-source models and decentralized development, heightens the risk of malicious use by both state and non-state actors, including terrorist groups and cyber mercenaries. Increasingly accessible dual-use and proprietary AI systems enable actors even with limited resources to gain access to capabilities that previously required significant investment or expertise, posing additional threats to international security and stability.
- *Algorithmic bias and ethical implications:* Algorithmic biases embedded within AI systems pose ethical and humanitarian concerns. Biases related to gender, race, age, or socioeconomic factors in AI datasets can intentionally and unintentionally perpetuate inequality and discrimination, particularly within sensitive military and security applications.
- *Digital divides and inequality:* Without deliberate policy actions, disparities between developed and developing nations in AI capabilities could deepen, increasing geopolitical tensions and socio-economic inequalities, thus undermining long-term global stability.

III. Relevant normative proposals

Microsoft recognizes several existing and emerging normative frameworks relevant to AI governance in the military domain, including:

- UNIDIR's RAISE initiative, facilitating international multi-stakeholder dialogues and governance proposals.
- The Responsible AI in the Military Domain (REAIM) Summits, emphasizing transparency, accountability, and human oversight at the international level.
- The US Department of Defense Responsible AI Strategy, highlighting responsibility, equitability, traceability, reliability, and governability.
- NATO's Principles of Responsible Use for AI in Defence, emphasizing reliability, governability, and traceability among member nations.

IV. Microsoft recommendations

To maximize opportunities and mitigate the challenges, Microsoft proposes several key recommendations:

- *Establish clear international norms and standards:* Develop explicit international norms and industry standards governing responsible use and development of military AI. These norms should delineate acceptable and unacceptable behaviors, providing robust frameworks to deter misuse and foster transparency and accountability, supported where appropriate by monitoring or compliance mechanisms. AI governance frameworks should explicitly differentiate operational contexts, such as peacekeeping, humanitarian assistance, crisis management, and conflict scenarios, to appropriately address varied ethical, legal, and humanitarian considerations. To ensure continued relevance, such norms should be periodically reviewed and updated to reflect evolving technological developments and operational realities.
- *Ensure human-centric oversight and accountability:* Adopt policies ensuring meaningful human judgment, oversight, and accountability remain central to military decisions involving AI, particularly regarding the use of force. Clear oversight mechanisms and enforceable accountability structures, including rigorous human control and review processes, are necessary to maintain ethical standards, avoid automation bias, and mitigate unintended consequences.
- *Advance secure and transparent AI development practices:* Promote rigorous technical standards and comprehensive life cycle management protocols covering pre-design, development, testing, deployment, operation, acquisition, and decommissioning. Robust vulnerability management, security audits, and transparent development and deployment processes should be integral components, alongside clear capacity-building measures, ensuring AI systems remain secure, responsible, and resilient throughout their operational life cycle.
- *Enhance responsible data governance practices:* Establish clear international guidelines on responsible data governance specifically tailored to military AI applications. Transparent and accountable data management practices addressing collection, sharing, storage, training, and operational usage are crucial for managing dual-use risks, preventing misuse, and maintaining strict compliance with international legal and ethical frameworks.
- *Address and reduce algorithmic bias:* Prioritize addressing algorithmic bias through rigorous testing, transparent data practices, and inclusive AI development processes. Developers and users should establish clear policies to proactively identify, mitigate, and remediate biases, especially when AI systems are deployed in sensitive military or security contexts.
- *Promote responsible innovation and risk-based regulation:* Support regulatory frameworks that are risk-based, outcome-oriented, and balanced, ensuring they encourage innovation while adequately addressing security and ethical risks associated with AI deployment. Industry should advocate for flexible, adaptive regulations that keep pace with technological change, without imposing overly prescriptive or impractical requirements. Industry-led initiatives, such as voluntary codes of conduct, vulnerability disclosure standards, and collaborative red-teaming exercises, should be actively supported and integrated into broader international normative frameworks.
- *Strengthen international governance and alignment:* Support and actively engage in international initiatives, including REAIM Summits and dialogues at the UN General Assembly and UN Security Council. Robust international governance frameworks, characterized by transparency, clear accountability measures, and trust-building mechanisms, are essential for coherent and inclusive approaches to AI governance. Member States and stakeholders should coordinate closely through these forums to reduce fragmentation and ensure global alignment.

-
- *Support knowledge-sharing and awareness-raising with the UN system:* Encourage and actively contribute to efforts by the UN Secretariat and relevant UN entities to convene meaningful multi-stakeholder expert dialogues, workshops, and knowledge-sharing on AI in the military domain. Exchanges through voluntary contributions, technical expertise, and collaborative initiatives should aim at enhancing global understanding of AI's implications for international peace and security.
 - *Strengthen international cooperation and information sharing:* Encourage robust international cooperation, emphasizing real-time threat intelligence sharing and joint attribution mechanisms. Industry actors should actively participate in collective cybersecurity efforts, enhancing global cybersecurity preparedness and response.
 - *Foster multi-stakeholder dialogue and collaboration:* Actively participate in and support forums such as RAISE, involving states, international organizations, academia, civil society, and industry. Such inclusive dialogues are essential for mutual understanding, shaping responsible AI practices, and developing collaborative governance structures.

V. Conclusion

Microsoft is deeply committed to proactive collaboration with Member States, the UN system, industry, and civil society to implement these recommendations swiftly and effectively. Through sustained collective efforts and ongoing engagement in multi-stakeholder initiatives, Microsoft will continue supporting responsible AI governance, innovation, and practices that meaningfully contribute to international peace and security.
