

1,4

# Inteligencia Artificial y Armas Nucleares

Por el CR(R) Dr OIM Osvaldo Azpitarte (\*)

## Resumen

La Inteligencia Artificial (IA) tiene un uso cada vez más generalizado, y su aplicación se ha extendido a casi todos los ámbitos del quehacer humano.

En el ámbito militar, el uso de la IA tiene un enorme potencial, y sus aplicaciones pueden emplearse para hacer mucho más rápido, eficiente y preciso el uso de los sistemas de armas, de los sistemas de inteligencia, vigilancia y reconocimiento, de los sistemas de comando, control y comunicaciones, y de los procesos de toma de decisiones.

El ámbito específico de las armas nucleares es particularmente influenciado por la aplicación creciente de IA. Aún desde la irrupción de los misiles intercontinentales (ICBM) en la década de 1960, los sistemas de armas nucleares y sus sistemas complementarios siempre se han caracterizado por su complejidad técnica, su automatismo y su autonomía, limitando la intervención humana a la toma de decisiones. Estas propiedades de los sistemas de armas nucleares pueden ser mejoradas exponencialmente por el uso de IA.

La aplicación de IA en los sistemas de armas nucleares, tiene, como presentaremos en este artículo, sus grandes ventajas, pero también sus desventajas. Estas desventajas deben ser tenidas cuidadosamente en cuenta, ya que ignorarlas puede conducir a graves consecuencias, teniendo en cuenta el enorme poder destructivo y devastador de las armas nucleares.

## Introducción

El término Inteligencia Artificial (IA) se refiere a la capacidad de algunos sistemas informáticos avanzados (combinación de software y hardware) de llevar a cabo tareas que normalmente requieren la intervención de inteligencia humana, entendida como la capacidad de reconocer, analizar y resolver problemas.

Todos los sistemas de IA combinan algoritmos de aprendizaje autónomo con cientos de capas de redes neuronales que pueden manejar billones de datos en forma simultánea. En cierta forma, estos sistemas constituyen una caja negra que recibe datos o inputs y provee respuestas o outputs.

Los sistemas de IA pueden ser utilizados para evaluar una enorme cantidad de datos de inteligencia de forma exponencialmente más rápida que un analista humano, y sin muchos de sus aspectos desfavorables, como ser el sesgo, la ira, el temor o el prejuicio, mejorando, así, los procesos de toma de decisiones. También puede responder a potenciales amenazas de forma más rápida y precisa que los operadores humanos.

Es importante reconocer que los seres humanos son influenciados, para la toma de decisiones, por su intuición, emociones y sentimientos, atributos que la IA no posee. La IA toma decisiones de forma diferente que los humanos, lo que puede ser tanto positivo como negativo. Por ejemplo, la IA puede tomar decisiones más rápido y sin influencia de las emociones, pero, a su vez, puede ignorar factores importantes que un humano sí puede considerar, como ser la experiencia o la apreciación de un contexto general.

La IA tiene un gran potencial en aplicaciones relacionadas con los sistemas de armas nucleares. Su empleo puede:

- > aumentar las capacidades de alerta temprana, inteligencia, vigilancia y reconocimiento,
- > acelerar los procesos de toma de decisiones,
- > mejorar la integración de los sistemas de comando, control y comunicaciones (NC3),
- > aumentar la capacidad de las armas nucleares autónomas, mejorando la maniobrabilidad, la capacidad de detección de obstáculos, y la identificación automática de blancos.

El uso de IA para armas nucleares lleva implícito un debate sobre tres áreas principales: la autonomía de las armas nucleares, la estabilidad de los sistemas militares de IA, y la estabilidad estratégica. Este debate se manifiesta en, entre otras, las siguientes consideraciones:

- > El uso de IA facilita el empleo de armas nucleares autónomas, entendiéndose por tales, aquellas que seleccionan el blanco y lo alcanzan sin intervención humana.
- > Los sistemas de IA carecen de la capacidad de analizar el contexto general de un conflicto, por lo que pueden provocar la toma de decisiones apresuradas. Los seres humanos son emocionales y pueden cometer errores, pero su experiencia y conocimiento del contexto puede compensar errores técnicos, como quedó demostrado en varios incidentes reales. Valga como ejemplo lo sucedido en 1983, cuando el sistema de alerta temprana de la entonces Unión Soviética indicó el lanzamiento de cinco misiles intercontinentales desde bases de USA. Stanislav Petrov, el oficial de servicio en el centro de comando del sistema de alerta temprana, sospechó que estas advertencias de ataques con misiles eran falsas alarmas, dada la escasa cantidad de misiles en vuelo, y decidió esperar a que se corroborara la evidencia, en lugar de transmitir inmediatamente la advertencia a la cadena de mando. La evidencia no se corroboró, y las investigaciones posteriores indicaron que hubo una falla técnica en el sistema de alerta. La prudente decisión del oficial Petrov evitó un ataque nuclear de represalia contra USA, lo que hubiera resultado en una guerra nuclear de gran escala.
- > El uso de IA facilita el empleo de los denominados “*dead-hand systems*”, sistemas que provocan una represalia automática, sin intervención humana, en caso de recibir un primer ataque nuclear o “*nuclear first strike*”.
- > Simulaciones hechas en juegos de guerra han demostrado que el uso de IA normalmente lleva a escalar un conflicto hasta llegar, incluso, al uso de armas nucleares.

## Marco legal

El documento internacional vigente más importante sobre el tema es la Declaración Política sobre el Uso Militar Responsable de la Inteligencia Artificial y la Autonomía, propuesto por el Departamento de Estado de USA, acordado en La Haya en 2023 y firmado por más de 50 países (Argentina aún no lo ha firmado).

Este documento define:

- > **Inteligencia Artificial (IA):** puede ser entendida como referida a la capacidad de las máqui-

nas de llevar a cabo tareas que hubieran, de otro modo, requerido inteligencia humana. Esto puede incluir: reconocimiento de patrones, aprendizaje de la experiencia, obtención de conclusiones, establecimiento de predicciones o generación de recomendaciones.

- > **Autonomía:** involucra sistemas que pueden operar sin intervención humana luego de su activación
- > **Capacidades de la IA militar:** incluye no solo a las armas si no también a los sistemas de apoyo a toma de decisiones que ayuden a los jefes militares de todos los niveles a tomar mejores y más oportunas decisiones, tanto en el campo de batalla como en el puesto de comando.

## El debate sobre la IA en la armas nucleares

### El potencial de la IA militar

La IA tiene un enorme potencial para aplicaciones militares. Las potencias nucleares la consideran esencial y buscan lograr su integración en los sistemas de armas, comando y control antes de que lo hagan sus adversarios. Se puede decir que se ha establecido una especie de carrera armamentista de IA (*AI arms race*).

La IA puede ahorrar tiempo y costo en la investigación, optimización del diseño, fabricación, prueba, mantenimiento y vigilancia de los misiles nucleares.

La IA también puede mejorar la alerta temprana y las capacidades de inteligencia, vigilancia y reconocimiento, coordinando el sensado, adquisición, procesamiento y diseminación de información relevante, precisa y oportuna, haciendo más eficiente y rápido el proceso de toma de decisiones.

### Autonomía de los sistemas de armas nucleares

Teniendo en cuenta el amplio poder devastador de las armas nucleares, este punto reviste particular importancia.

Una clasificación ampliamente aceptada de la intervención humana en los sistemas autónomos es la siguiente:

- > *“Human in the loop”*: sistemas de armas que usan su autonomía para atacar blancos que un humano debe decidir atacar.
- > *“Human on the loop”*: sistemas de armas que usan su autonomía para atacar blancos, pero los controladores humanos pueden detener su operación cuando sea necesario.
- > *“Human out of the loop”*: sistemas de armas que usan su autonomía para atacar blancos específicos sin posible intervención de operadores humanos.

La opción más ampliamente aceptada por las potencias nucleares es la de mantener los sistemas de armas nucleares autónomos como *“Human in or on the loop”*.

### La estabilidad de la IA militar

Los sistemas militares autónomos pueden mejorar el reconocimiento y la inteligencia, acelerar el proceso de toma de decisiones y permitir reacciones más rápidas, pero pueden, también, desestabilizar o perjudicar el cumplimiento de una misión militar. Algunos ejemplos pueden servir para ilustrar:

- > Un sistema autónomo puede decidir atacar un blanco importante, pero esto puede revelar o poner en riesgo una operación encubierta.

- > Una computadora puede considerar que combatir por la conquista de una localidad puede significar un desperdicio de recursos, y decidir, en consecuencia, una retirada, sin tener en cuenta que esa localidad puede tener un alto valor simbólico o psicológico.
- > La IA puede decidir realizar un ataque en forma apresurada, sin tener en cuenta la necesaria preparación de las fuerzas convencionales, o cerrando la puerta a posibles soluciones pacíficas. Se ha comprobado que todos los modelos de IA tienden a adoptar, siempre, actitudes agresivas, que incluyen el uso de armas nucleares.

## Vulnerabilidades o fallas de los sistemas de IA

Cuando se habla de falla de los sistemas de IA se debe entender que no es que los sistemas de IA funcionen mal. Los sistemas hacen exactamente aquello para lo que fueron desarrollados. Lo que a veces sucede, es que esto no siempre concuerda con el deseo de los desarrolladores.

Los sistemas de IA pueden presentar las siguientes vulnerabilidades:

- > Envenenamiento de datos: los sistemas de IA reciben, de forma casi instantánea, millones de datos provenientes de diferentes fuentes, pero carecen de la capacidad de distinguir la confiabilidad de esos datos. La manipulación de esos datos, deliberada o no, puede resultar en resultados indeseables, que, tratándose de armas nucleares, pueden tener graves consecuencias (por ejemplo, malinterpretar un ataque nuclear entrante).
- > Manipulación de imágenes: la manipulación de imágenes digitales, deliberada o no, puede provocar, entre otros efectos, una mala interpretación por parte de los sistemas de vigilancia y reconocimiento.
- > Sesgo automático: producto del proceso de aprendizaje autónomo, los sistemas de IA pueden tener una tendencia, que puede ser incluso imperceptible, a juzgar los hechos en una forma no del todo ecuánime, lo que en el caso de un conflicto nuclear de rápido desarrollo puede tener graves consecuencias.
- > Escalada artificial: cuando un sistema de IA realiza cálculos basados en el output de otro sistema de IA, se puede producir una realimentación positiva que escale un conflicto de forma artificial.
- > Ataques cibernéticos: como cualquier otro software, los sistemas de IA pueden sufrir todo tipo de ataque cibernético, sobre todo de sistemas de IA enemigos. Estos ataques pueden incluir malware, virus, ransomware, etc..

Todas estas vulnerabilidades, hacen que, en general, a los sistemas de IA no se les permita adoptar una decisión final, sobre todo cuando se trata del uso de armas nucleares.

## La IA y la estabilidad estratégica

El mantenimiento de la estabilidad estratégica implica evitar, en lo posible, cualquier evento que dé lugar a un primer ataque nuclear. La política conocida como "primer ataque nuclear" (*first strike policy*), implica atacar primero el arsenal nuclear enemigo para evitar que éste lance una represalia, o para minimizarla.

Una razón para lanzar un primer ataque nuclear puede ser la incerteza sobre las capacidades nucleares del enemigo, ya que si se ignora lo que éste puede hacer ante determinada situación, la única forma de asegurar una victoria es atacar primero. En este sentido, la IA puede contribuir a presentar un conflicto nuclear como "ganable", ya que puede socavar la disuasión nuclear del enemigo amenazando su capacidad de responder a un primer ataque.

La IA también puede perjudicar la estabilidad estratégica por la reducción de los tiempos para la toma de decisiones, lo que puede resultar en una escalada del conflicto o en un uso involuntario de armas nucleares.

Tanto la aparición de las armas hipersónicas, cuya detección oportuna es muy difícil, como la creciente velocidad de desarrollo de los conflictos, socavan la estabilidad estratégica e incrementan el riesgo de un enfrentamiento nuclear.

## Recomendaciones para la aplicación de IA a las armas nucleares

Algunos expertos sugieren la adopción de las siguientes medidas de seguridad para el empleo de IA en armas nucleares:

- > Mejorar la seguridad de las armas nucleares mediante el aumento de salvaguardas y la evaluación de riesgos de ataques cibernéticos.
- > Implementar códigos de autenticación para mejorar los protocolos de comando y control y los mecanismos para disminuir la probabilidad de un lanzamiento nuclear involuntario.
- > Utilizar salvaguardas robustas para limitar las consecuencias de errores o accidentes
- > Establecer acuerdos bilaterales y multilaterales para fortalecer la confianza mutua e incrementar el diálogo sobre el uso de la IA en el control de las armas nucleares.

## Conclusiones

En el ámbito de las armas nucleares, el uso de IA es considerado como positivo para funciones básicas como comunicaciones, diseño, pruebas, etc.. mientras que hay recelo general en lo que concierne a incluirla en procesos de toma de decisiones y lanzamientos autónomos. La imprevisibilidad y susceptibilidad a ataques cibernéticos son, por ahora, los argumentos que impiden la inclusión de la IA en los procesos de toma de decisiones.

Existe actualmente, entre las potencias nucleares, un consenso sobre el que la IA es una tecnología aún inmadura como para ser utilizada en escenarios estratégicos de alto riesgo, y que su uso debería limitarse al apoyo de las funciones de comando y control. En ese sentido, en mayo de 2024, el Departamento de Estado de USA pidió a China y Rusia que declaren oficialmente que no darán el control de sus armas nucleares a los sistemas de IA, como ya fuera comprometido con el Reino Unido y Francia en 2022.

## Referencias

- [1] Artificial Intelligence and Nuclear Weapons, Klauss Saalbach, Geostrategy and Geopolitics Department, Osnabrück University, Germany, 2024.
- [2] Stanislav Petrov: el hombre que salvó al mundo de un desastre nuclear, BBC News, [https://www.bbc.com/mundo/noticias/2013/09/130926\\_internacional\\_ruso\\_detuvo\\_ataque\\_nuclear\\_jrg](https://www.bbc.com/mundo/noticias/2013/09/130926_internacional_ruso_detuvo_ataque_nuclear_jrg), 2013.
- [3] Dr James Johnson on How AI is Transforming Nuclear Deterrence, Caroline Russell, NTI – Nuclear Threat Initiative, 2023.
- [4] Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, U.S. Department of State, 2023.
- [5] On Integrating Artificial Intelligence with Nuclear Control, Peter Rautenbach, Arms Control Association, 2022.
- [6] Artificial Intelligence and the Evolving Landscape of Nuclear Strategy, Silky Kaur, Union of Concerned Scientists, 2024.
- [7] AI and Nuclear Weapons: Keeping the human in the loop, not only for the decision, but also before the decision, Paul van Hooff, The Hague Centre for Strategic Studies, 2024.

**(\*) Osvaldo Azpitarte** es Coronel retirado de Artillería, promoción 107 del Colegio Militar de la Nación. Egresado de la Escuela Superior Técnica como Ingeniero Químico, en 1993. Egresado del Instituto Balseiro (CNEA), como Ingeniero Nuclear, en 1993. Doctor en Ingeniería Nuclear, por el Instituto Balseiro, en el año 2003.