ACADEMIA DE LAS CIENCIAS
Y LAS ARTES MILITARES

Comunicaciones académicas

# Aspectos éticos del uso de la inteligencia artificial en operaciones militares en el dominio cognitivo

Manuel Esteve Domingo
Academia de las Ciencias y las Artes Militares
Sección de Futuro de las Operaciones Militares

21 de febrero de 2026

## Introducción

Los pilares de la IA ética en cualquier aplicación y dominio son la equidad, la responsabilidad y la transparencia. El objetivo es crear un proceso algorítmico que no genere sesgos, discriminación ni consecuencias injustas para sus usuarios, que muestre claramente la responsabilidad de las acciones y además sea explicable. Los sistemas basados en IA deben ser seguros, justos, explicables, responsables, controlables y respetuosos con los derechos y la dignidad humana.

El uso de la IA no está exento de riesgos y limitaciones como son los errores en los algoritmos, sesgo en los datos, falta de explicabilidad, y posibles amenazas para la privacidad y seguridad de datos personales. Particularmente, la IA generativa presenta riesgos y limitaciones adicionales como el sesgo algorítmico, las alucinaciones que pueden producir información inventada sin base real debido a interpretaciones erróneas de patrones o falta de datos precisos, la creación de contenidos deliberadamente falsos, y muy grave, la sobreconfianza de los usuarios en los resultados generados.

Cuando nos centramos en el uso de la IA en el ámbito militar tenemos que tener en cuenta los principales marcos de referencia que regulan su uso, principalmente el definido por OTAN, denominado *Principios de Uso Responsable de IA en Defensa*, que incide en la legalidad, responsabilidad, rendición de cuentas, explicabilidad y trazabilidad, fiabilidad, gobernabilidad y mitigación de sesgos. Muy similar al del DoD de EE. UU., que incide en que el uso de la IA debe ser responsable, equitativo, trazable, fiable y gobernable.

## Ética de la IA en Defensa

La ética de la IA en Defensa es un marco muy amplio que incluye distintos ámbitos de aplicación como logística, inteligencia, ciberdefensa y, de forma muy extendida, apoyo a la toma de decisiones en todos los niveles de mando y control. Se puede clasificar la aplicación de la IA según cercanía a la fuerza: back-office (logística, mantenimiento), ISR/Análisis (detección, clasificación), apoyo a la decisión (recomendaciones de objetivos, priorización), control de plataforma (navegación, seguimiento), y, por último, selección y ataque.

Para analizar el uso ético de la IA en un sistema de aplicación militar debemos preguntarnos:

1. ¿El sistema selecciona objetivos?
2. ¿El sistema decide establecer enfrentamiento o atacar de forma autónoma?
3. ¿Puede hacerlo sin intervención humana adicional una vez activado?

Evidentemente, cuanto más cercano sea el uso de la IA al enfrentamiento físico, los aspectos éticos son más relevantes y su análisis previo más necesario.

## Aplicaciones militares de la IA en el dominio cognitivo

Las operaciones de influencia en el dominio cognitivo persiguen producir efectos, tanto en el propio dominio cognitivo como en el resto de dominios. Cuando estas operaciones están reforzadas por el uso de IA, todo el proceso, tanto de planeamiento como de conducción, así como los efectos que se consiguen, se aceleran y se amplifican. Los aspectos éticos del uso de la IA, por tanto, deben analizarse como el uso de cualquier otra capacidad ofensiva.

La IA aporta a las operaciones en el dominio cognitivo capacidades típicas como: análisis de audiencias, segmentación, detección de tendencias, modelado de sentimientos, generación de contenido a gran escala (texto, audio, video), automatización de la difusión de contenidos sintéticos (bots, cuentas coordinadas, respuesta «en tiempo real»), así como evaluación de los efectos, optimización y ajuste de mensajes.

En el dominio cognitivo, esas capacidades que aporta la IA aumentan los riesgos éticos como son la manipulación, la distorsión de la capacidad de decidir informadamente, el engaño y la suplantación *(deepfakes*, «pruebas» fabricadas, falsas atribuciones). Pero todavía son más graves los riesgos como la explotación de vulnerabilidades de grupos sensibles por edad, discapacidad o situación social o económica, así como la generación de daños sociales, polarización, pánico, violencia y erosión de la confianza pública, incluida la confianza en instituciones.

Los principios éticos aplicables en el uso de la IA en el dominio cognitivo son la legalidad y legitimidad, la no manipulación y no explotación de vulnerabilidades, la transparencia y trazabilidad, la responsabilidad y capacidad de rendición de cuentas y, en definitiva, la minimización del daño conocido como proporcionalidad «cognitiva», principalmente sobre la población civil no combatiente.

Existen algunos límites éticamente muy difíciles de sobrepasar como son la generación de *deepfakes* o contenidos falsos para engañar a la población civil, aunque sea con el objetivo de obtener algún tipo de ventaja militar. Igualmente, la desinformación deliberada dirigida a públicos vulnerables, con el objetivo de explotar vulnerabilidades psicológicas o socioeconómicas, y la microsegmentación de las audiencias basada en datos sensibles como salud o religión para explotar miedos o fragilidades, se debe considerar como más allá de lo ético.

Las técnicas de manipulación subliminal o persuasión que anulen la decisión informada están explícitamente restringidas en la legislación de la UE, así como industrializar mensajes que incentiven el odio, la violencia o la degradación de la dignidad humana. El uso ético de la IA en operaciones militares en el dominio cognitivo es una zona especialmente sensible: no se trata solo de hacer un uso responsable de la IA, teniendo en cuenta sesgos, seguridad y trazabilidad, sino de impactar en percepciones, creencias y decisiones humanas, a veces de poblaciones civiles, con herramientas que pueden escalar, personalizar y automatizar la persuasión, hasta producir daños letales en la población civil. En ausencia de una estricta supervisión, el uso de la IA en estas operaciones puede ser asimilable al uso de armas letales autónomas.

## Relación entre el uso de la IA en el dominio cognitivo y en las armas autónomas letales

Las armas letales autónomas es un campo donde se ha analizado mucho el uso ético de la IA. La relación ética es que en ambos casos la IA actúa sobre decisiones humanas. En influencia cognitiva, la IA puede alterar percepciones, creencias y voluntades. En un paso más allá, en las armas autónomas, la IA puede acelerar o tomar decisiones de vida o muerte; por eso comparten un núcleo ético común (derechos, responsabilidad, control humano, trazabilidad), pero con umbrales y salvaguardas más estrictos cuando hay uso de la fuerza. Las armas autónomas son un caso extremo (y de alto riesgo) de aplicación de IA. Concentran casi todos los dilemas de la IA (sesgos, opacidad, responsabilidad, seguridad) y además añaden uno decisivo: la decisión de causar muerte.

Tanto en las operaciones en el dominio cognitivo como en el uso de armas letales autónomas, se debe tener en cuenta la ética y la ley del conflicto armado (Derecho Internacional Humanitario), respetando tres principios fundamentales:

1. Distinción: distinguir combatientes de civiles.
2. Proporcionalidad: evitar daño civil excesivo respecto a la ventaja militar.
3. Precaución: tomar medidas para minimizar el daño.

La ética aplicada en el uso de armas letales autónomas puede ilustrarnos sobre cómo aplicar principios éticos en las operaciones en el dominio cognitivo. Algunos de estos principios son el control humano significativo y efectivo (no solo nominal), la supervisión real, la comprensión del contexto, la posibilidad de abortar y la gestión de fallo seguro: si el sistema se degrada o hay incertidumbre, debe «irse a seguro» y no escalar la letalidad. Por otra parte, se debe exigir al uso de la IA trazabilidad reforzada, *logs* forenses y revisiones legales del uso de las armas. Sin trazabilidad, la rendición de cuentas se vuelve muy difícil, si no imposible.

Por otra parte, la ética de la IA aplicada en las operaciones en el dominio cognitivo debe respetar los siguientes principios, no muy distintos a los requeridos en las armas letales autónomas. En primer lugar, deben existir límites de objetivo y proporcionalidad. No todo lo eficaz es ético y se debe excluir prácticas que busquen deliberadamente dañar a civiles o anular su autonomía moral como campañas de pánico o acoso. Un segundo principio es el de transparencia y no engaño, según el contexto: en entornos no bélicos, el uso de IA para suplantación y manipulación masiva es difícil de justificar éticamente. En conflicto, incluso si hay engaño permitido, siguen existiendo líneas rojas legales. Y finalmente el control de daños, que requiere una evaluación previa del riesgo de efectos secundarios (polarización, violencia, persecución), y mecanismos para detener o rectificar la acción cognitiva.

Tras los análisis previos, podemos definir unos principios éticos comunes en el uso de la IA aplicables a las operaciones en el dominio cognitivo y a las armas letales autónomas:

1. Responsabilidad (*Accountability*): siempre debe poder identificarse quién decidió, autorizó, desplegó y con qué reglas.
2. Trazabilidad y explicabilidad suficiente: registrar datos, modelos, *prompts*, versiones, reglas de decisión y evidencias usadas, especialmente si afectan a derechos o al uso de la fuerza.
3. Gobernabilidad: capacidad real de detectar comportamientos no deseados y capacidad para desactivar o abortar la operación.
4. Equidad y minimización de sesgos: evitar que la IA amplifique discriminación (poblaciones, idiomas, perfiles) o genere «objetivos» por proxies.
5. Fiabilidad y seguridad: robustez ante manipulación por parte del adversario, ciberseguridad y rendimiento verificable en el contexto real de uso.

## Principios éticos básicos en la aplicación de IA en el dominio cognitivo

La pregunta ética clave sería: ¿la operación persigue un fin legítimo (seguridad/defensa) y está autorizada por el marco legal aplicable?

En entornos democráticos, esto incluye no solo la ley militar, sino derechos fundamentales (libertad de expresión, privacidad, no discriminación) y controles institucionales. Marcos internacionales de ética de IA ponen el énfasis en los derechos humanos y la dignidad como base. La *AI Act* de la UE prohíbe ciertas prácticas de IA que usan técnicas subliminales o manipulativas que distorsionan materialmente el comportamiento causando daño, y también prohíbe explotar vulnerabilidades de grupos, entre otras prácticas.

Por otra parte, en operaciones cognitivas la transparencia no es solo una cuestión estética, debe ser el antídoto contra la suplantación. En la UE, el marco regulatorio obliga a marcar o etiquetar contenidos generados o manipulados por IA y a reducir riesgos de engaño, *impersonación* y desinformación, a través de un *Code of Practice* en desarrollo ligado a obligaciones de transparencia

Otro principio fundamental es la responsabilidad y rendición de cuentas, y una regla práctica es que, si no se puede auditar y explicar qué se hizo, no se debería hacer con IA automatizada. Este enfoque conecta bien con principios de IA responsable en Defensa como los de la OTAN: legalidad, responsabilidad, explicabilidad, trazabilidad, fiabilidad, gobernabilidad y mitigación de sesgos.

La ética exige calibrar necesidad y proporcionalidad con el objetivo de minimización del daño, en lo que se conoce como proporcionalidad «cognitiva». Aunque las operaciones en el dominio cognitivo no se consideren letales en sí mismas, sí que pueden producir daños reales incluso en el dominio físico, como son la desconfianza social, la estigmatización de grupos, los efectos psicológicos, la erosión democrática y la violencia inducida.

Por ello se requiere realizar una evaluación de impacto previa que debe tener en cuenta los siguientes aspectos: propósito legítimo y necesidad, audiencias (¿combatientes? ¿civiles? ¿población propia?), nivel de automatización, riesgos previsibles (daño social, violencia, estigmatización), mitigaciones y criterios de parada.

Todo ello nos lleva al concepto *Human-in-command* real. Un humano responsable debe aprobar el contenido de la operación y su despliegue cuando hay riesgo alto, sobre todo para la población no combatiente. En cualquier caso, se debe contemplar la prohibición de la publicación autónoma de contenidos sintéticos generados por IA en situaciones de crisis sin revisión previa por un humano, en conexión con los principios de responsabilidad y gobernabilidad definidos en marcos de uso de la IA en Defensa (OTAN/DoD).

Y ello nos conduce al denominado modelo «Centauro», mitad hombre, mitad IA.

## Conclusiones

Las operaciones militares en el dominio cognitivo deben tener en cuenta la ética habitual del uso de la IA y, además, aplicar salvaguardas reforzadas por su impacto directo en personas y decisiones.

Los principios éticos de la IA deben aplicarse de forma vinculante, preventiva y verificable, garantizando control humano significativo, legalidad y rendición de cuentas, proporcionalidad y minimización de daños, y transparencia y capacidad de auditoría para evitar manipulación indebida y proteger, sobre todo, a los civiles y sus derechos.

La IA en operaciones militares en el dominio cognitivo debe regirse por consideraciones éticas comparables a las de las armas letales autónomas (control humano significativo, proporcionalidad, minimización del daño y rendición de cuentas), porque también se pueden causar daños graves, principalmente en la población civil, aunque no siempre sean daños físicos. ▮