



15/10/2026

Doble uso, opacidad y disuasión: la IA y la erosión de la señalización estratégica

Dual use, opacity and deterrence: artificial intelligence and the erosion of strategic signalling

Abstract:

The stability of the international system rests partly on signalling mechanisms that allow states to communicate their military capabilities and reduce the risk of conflict through miscalculation. Artificial intelligence (AI) changes the conditions of this exchange. Its dual-use character, the opacity of certain models and the difficulty of inferring operational capabilities from observable infrastructure weaken some foundations of nuclear and conventional signalling. This article examines the implications for deterrence, cyber attribution and international verification, without equating opacity with the impossibility of oversight. It argues that AI can improve coordination across domains, although such coordination does not necessarily depend on it. Finally, it assesses the possibilities and limits of the European framework, distinguishes civilian regulation from exclusively military uses and proposes developing transparency and confidence-building mechanisms compatible with the protection of sensitive information.

Keywords:

Artificial intelligence; dual use; strategic signalling; deterrence; verification.

Cómo citar este documento:

MUÑOZ PAIRET, Pau. *Doble uso, opacidad y disuasión: la inteligencia artificial y la erosión de la señalización estratégica*. Documento de Opinión IEEE 103/2026.

Introducción

Los recursos naturales, demográficos y tecnológicos no están distribuidos de manera uniforme entre los Estados. Desde una lectura realista, esta asimetría, unida a la ausencia de una autoridad supranacional efectiva, alimenta la competencia por la seguridad. Los Estados procuran preservar su supervivencia mediante el poder militar, entendido tanto como capacidad para ejercer fuerza hacia el exterior como para defenderse de amenazas de terceros y proteger su población, territorio e instituciones¹. La superioridad relativa puede facilitar la coerción, pero no convierte la guerra en una consecuencia inevitable: también sirve para sostener la defensa, la negociación y la disuasión.

En este contexto, los gobiernos emplean mecanismos de señalización para comunicar capacidades e intenciones y para interpretar las de sus adversarios. Los trabajos de Thomas Schelling y Robert Jervis permiten entender cómo las amenazas, los compromisos y las percepciones influyen en el cálculo estratégico². Los ejercicios, las adquisiciones de armamento y los ensayos tecnológicos proporcionan información, aunque su significado depende del contexto político y de la credibilidad de quien los realiza.

La irrupción de la inteligencia artificial (IA), como tecnología transversal de doble uso, modifica esa arquitectura. Un sistema puede cambiar de comportamiento mediante ajustes en sus modelos, datos o software sin que su apariencia exterior revele la transformación. La tesis de este artículo es que esa distancia entre lo observable y la capacidad efectiva erosiona determinados fundamentos de la señalización estratégica y exige adaptar los instrumentos de confianza y verificación.

El análisis es conceptual y se apoya en bibliografía estratégica, documentos institucionales y ejemplos públicos. No pretende medir el grado de opacidad de todos los sistemas de IA ni demostrar una relación causal uniforme entre su adopción y el conflicto. Su propósito es identificar mecanismos de incertidumbre y las condiciones bajo las cuales pueden afectar a la credibilidad de las señales.

¹ MEARSHEIMER, John J. The Tragedy of Great Power Politics. W. W. Norton, Nueva York, 2001.

² SCHELLING, Thomas C. The Strategy of Conflict. Harvard University Press, Cambridge (MA), 1960. SCHELLING, Thomas C. Arms and Influence. Yale University Press, New Haven, 1966. JERVIS, Robert. Perception and Misperception in International Politics. Princeton University Press, Princeton, 1976.

El sistema internacional como dinámica de la información

En el realismo ofensivo, la búsqueda de poder relativo ocupa un lugar central en la explicación del comportamiento de las grandes potencias³. Sin embargo, la inseguridad también puede conducir a estrategias defensivas: un Estado puede reforzar sus capacidades para resistir una agresión, mientras su adversario interpreta ese refuerzo como preparación ofensiva. El problema no reside únicamente en la distribución material del poder, sino en la información disponible sobre su finalidad.

Los Estados se enfrentan así a preguntas recurrentes: ¿cuál es la fuerza real del adversario? ¿Hasta qué punto es creíble su amenaza o su disposición a aceptar un acuerdo? La teoría de la negociación estratégica muestra que comunicar capacidades y determinación puede modificar los incentivos sin recurrir al empleo efectivo de la fuerza⁴. Una señal útil debe ofrecer información relevante y resultar suficientemente costosa, contrastable o vinculante para distinguirse de una mera declaración.

Cuando la señalización funciona, las partes ajustan sus expectativas y amplían el espacio de negociación. Cuando falla, por engaño, errores de interpretación o dificultades de verificación, crece el riesgo de cálculo equivocado. La información no explica por sí sola el estallido de una guerra: intervienen intereses incompatibles, problemas de compromiso y decisiones políticas. Su calidad sí condiciona la posibilidad de anticipar costes, reconocer límites y evitar escaladas involuntarias. La estabilidad depende, por tanto, de capacidades defensivas creíbles y de una comunicación que permita interpretarlas.

También importa distinguir entre informar sobre una capacidad y comprometerse a utilizarla. Un ejercicio puede demostrar cierta preparación sin revelar qué decisión tomará un gobierno bajo presión. A la inversa, una declaración política es capaz de expresar determinación sin acreditar los medios disponibles. La interpretación estratégica combina ambas dimensiones y las contrasta con la conducta observada. Esta distinción evita atribuir a la transparencia técnica una función que depende igualmente de las relaciones políticas y de la confianza acumulada entre las partes.

³ MEARSHEIMER, John J. *The Tragedy of Great Power Politics*, op. cit.

⁴ SCHELLING, Thomas C. *The Strategy of Conflict y Arms and Influence*, op. cit.

La señalización estratégica clásica: del armamento nuclear a la disuasión convencional

El armamento nuclear ofrece un caso especialmente relevante de señalización. Su elevado coste, las instalaciones que requiere y determinadas huellas físicas de su desarrollo permiten obtener indicios externos de capacidad⁵. Los ensayos explosivos pueden detectarse mediante técnicas sísmicas, hidroacústicas, infrasónicas y de radionúclidos⁶. Esta observabilidad no equivale a transparencia completa: persisten incertidumbres sobre arsenales, doctrinas, disponibilidad y voluntad de empleo.

Sobre una parte de esas propiedades se construyeron instrumentos jurídicos y técnicos de verificación. El Tratado sobre la No Proliferación de las Armas Nucleares (TNP) establece obligaciones diferenciadas, y las salvaguardias del Organismo Internacional de Energía Atómica (OIEA) verifican el cumplimiento de compromisos relativos al uso pacífico de materiales y tecnologías nucleares⁷. Este sistema no debe confundirse con una inspección general de todas las cabezas nucleares o los vectores militares. Su eficacia depende de los acuerdos aplicables, del acceso disponible y de la cooperación estatal.

La experiencia nuclear muestra que es posible sostener la disuasión con información incompleta. La capacidad de represalia y la percepción de daños inaceptables han contribuido a contener enfrentamientos directos entre potencias nucleares, aunque no han eliminado las crisis ni las guerras por intermediarios. La estabilidad no procede de una tecnología perfectamente legible, sino de una combinación de capacidades, señales, instituciones y aprendizaje político⁸.

La carrera espacial ilustra que el doble uso tampoco es una novedad. El lanzamiento del Sputnik en 1957 fue un logro científico y una señal del desarrollo soviético de cohetes de largo alcance⁹. Su órbita era observable por terceros, aunque no demostraba por sí sola

⁵ BRODIE, Bernard. *Strategy in the Missile Age*. Princeton University Press, Princeton, 1959. FREEDMAN, Lawrence. *The Evolution of Nuclear Strategy*. 3.ª ed., Palgrave Macmillan, Basingstoke, 2003.

⁶ CTBTO. *Monitoring Technologies*. Disponible en: <https://www.ctbto.org/our-work/monitoring-technologies> (consulta: 03/10/2026).

⁷ NACIONES UNIDAS. *Tratado sobre la No Proliferación de las Armas Nucleares*. 1968, artículo III. Disponible en: <https://www.un.org/en/conf/npt/2010/npttext.shtml> (consulta: 03/10/2026). OIEA. *Nuclear safeguards explained*. Disponible en: <https://www.iaea.org/topics/safeguards-explained> (consulta: 03/10/2026).

⁸ FREEDMAN, Lawrence. *The Evolution of Nuclear Strategy*, op. cit.

⁹ NASA. *65 Years Ago: Sputnik Ushers in the Space Age*. 4 de octubre de 2022. Disponible en: <https://www.nasa.gov/history/65-years-ago-sputnik-ushers-in-the-space-age/> (consulta: 03/10/2026).

todas las prestaciones de un arsenal operativo. La distinción es importante: una prueba física puede acreditar una capacidad delimitada sin resolver todas las incertidumbres sobre su empleo militar.

El problema que plantea la IA debe formularse en términos comparativos. Muchas tecnologías combinan usos civiles y militares; lo relevante es qué relación cabe establecer entre la evidencia externa y una capacidad concreta. En los sistemas convencionales también existen ocultación y engaño. La diferencia aparece cuando una parte creciente de la capacidad depende de componentes informacionales que pueden modificarse sin cambios materiales equivalentes.

La inteligencia artificial como tecnología de doble uso: revisión de los fundamentos clásicos

La IA puede intervenir en el análisis de información, la logística, el apoyo a decisiones y el funcionamiento de plataformas con distintos grados de autonomía. Su efecto militar está condicionado por la integración técnica, la organización y la formación, además del rendimiento del algoritmo¹⁰. Tampoco deben identificarse automáticamente IA, automatización y autonomía: son conceptos relacionados, pero una plataforma no tripulada puede estar controlada a distancia y un sistema de apoyo puede incorporar IA sin decidir por sí mismo¹¹.

Los desarrollos industriales muestran esta diversidad. Saab anunció en junio de 2025 tres vuelos con el agente Centaur de Helsing integrado en un Gripen E; el primero se realizó el 28 de mayo¹². Helsing presentó en septiembre de 2025 el proyecto CA-1 Europa mediante una maqueta a escala real, cuya entrada en servicio sigue siendo un objetivo de desarrollo¹³. Por su parte, el Ejército de Tierra estadounidense anunció en

¹⁰ HOROWITZ, Michael C. Artificial Intelligence, International Competition, and the Balance of Power. Texas National Security Review, vol. 1, núm. 3, 2018, pp. 36–57. Disponible en: <https://doi.org/10.15781/T2639KP49> (consulta: 03/10/2026).

¹¹ BOULANIN, Vincent y VERBRUGGEN, Maaïke. Mapping the Development of Autonomy in Weapon Systems. SIPRI, Estocolmo, 2017. Disponible en: <https://www.sipri.org/publications/2017/policy-reports/mapping-development-autonomy-weapon-systems> (consulta: 03/10/2026). SCHARRE, Paul. Army of None: Autonomous Weapons and the Future of War. W. W. Norton, Nueva York, 2018.

¹² SAAB. Saab achieves AI milestone with Gripen E. 11 de junio de 2025. Disponible en: <https://www.saab.com/newsroom/press-releases/2025/saab-achieves-ai-milestone-with-gripen-e> (consulta: 03/10/2026).

¹³ HELSING. Helsing unveils CA-1 Europa, an autonomous fighter jet. 25 de septiembre de 2025. Disponible en: <https://helsing.ai/newsroom/helsing-unveils-ca-1-europa-an-autonomous-fighter-jet> (consulta: 03/10/2026).

marzo de 2026 un contrato marco con Anduril, de hasta veinte mil millones de dólares, para consolidar la adquisición de sus soluciones comerciales¹⁴. Las pruebas, los anuncios industriales y los techos contractuales constituyen señales diferentes; ninguno acredita por sí solo un despliegue completo ni un rendimiento en combate.

Esta evolución se relaciona con las operaciones multidominio. La OTAN distingue los dominios terrestre, marítimo, aéreo, espacial y cibernético, con efectos que también se proyectan sobre la dimensión cognitiva¹⁵. La IA puede facilitar la fusión de información y mejorar la coordinación entre dominios, pero no es una condición imprescindible: los ataques y las respuestas pueden planificarse y coordinarse mediante procedimientos tradicionales. La experiencia de Ucrania, analizada en documentos del Instituto Español de Estudios Estratégicos (IEEE), muestra la interacción de capacidades convencionales, digitales y no tripuladas, cuyo resultado depende asimismo del mando y de la adaptación organizativa¹⁶.

La dimensión informativa añade otra dificultad. Los modelos generativos pueden abaratar la producción de contenido sintético y facilitar la presentación engañosa de capacidades o acontecimientos¹⁷. Eso no elimina la posibilidad de contrastar una imagen o una declaración con sensores, fuentes independientes y análisis de contexto. Sí incrementa el esfuerzo de autenticación necesario para que las señales públicas resulten creíbles. El engaño precede a la IA; cambian su escala, velocidad y accesibilidad.

La infraestructura constituye un segundo foco de ambigüedad. Un centro de datos puede alojar servicios civiles y aplicaciones militares, aunque no todas las instalaciones tengan las prestaciones requeridas para entrenar modelos avanzados. Los chips, el suministro

¹⁴ U.S. ARMY PUBLIC AFFAIRS. U.S. Army awards enterprise contract for IT commercial solutions. 13 de marzo de 2026. Disponible en: https://www.army.mil/article/291074/u_s_army_awards_enterprise_contract_for_it_commercial_solutions (consulta: 03/10/2026).

¹⁵ NATO ALLIED COMMAND TRANSFORMATION. Multi-Domain Operations. Disponible en: <https://www.act.nato.int/activities/multi-domain-operations/> (consulta: 03/10/2026).

¹⁶ SAURÍN MARTÍNEZ, Daniel. Origen y fundamentos teóricos de la guerra multidominio y su aplicación en la guerra de Ucrania. Documento de Opinión IEEE 100/2025. Disponible en: https://www.defensa.gob.es/ceseden/-/ieeee/guerra_multidominio_2025_dieeee100 (consulta: 03/10/2026). ABAD SOTO, Jorge. Cambio de paradigma de la seguridad y la defensa: capacidades futuras, ya presentes. Documento de Opinión IEEE 04/2025. Disponible en: https://www.defensa.gob.es/ceseden/-/ieeee/cambio_de_paradigma_de_la_seguridad_y_la_defensa_2025_dieeee04 (consulta: 03/10/2026).

¹⁷ GARCÍA HUETE, Estíbaliz et al. IA generativa y sistemas serverless en desinformación bélica con LUTs: impacto y ética. Documento de Opinión IEEE 45/2025. Disponible en: https://www.defensa.gob.es/ceseden/-/ieeee/ia_generativa_y_sistemas_serverless_en_desinformacion_belica_con_luts_impacto_y_etica (consulta: 03/10/2026).

eléctrico y las instalaciones siguen siendo físicos y potencialmente observables; resulta más difícil inferir qué modelo se ejecuta, con qué datos y con qué finalidad¹⁸. La capacidad de cómputo ofrece un indicador parcial, no una medida directa de eficacia militar.

La naturaleza distribuida de la cadena de valor complica esa inferencia. Modelos, datos y aplicaciones circulan entre organizaciones, mientras la integración con sensores, comunicaciones y procedimientos operativos determina su utilidad efectiva. Por ello, contar equipos o inspeccionar instalaciones sigue aportando información, pero no basta para verificar todas las capacidades. La tabla 1 resume estas diferencias como tendencias analíticas, sujetas a variación según el sistema considerado.

Tabla 1. Señalización clásica y capacidades basadas en IA

Dimensión	Capacidades clásicas	Capacidades basadas en IA
Evidencia externa	Firmas físicas; acceso condicionado	Infraestructura observable; función menos visible
Localización	Instalaciones y plataformas identificables	Cómputo localizado; servicios distribuidos
Doble uso	Presente en componentes y vectores	Extendido en datos, modelos e infraestructura
Atribución	Variable según sistema y contexto	Compleja en ciberespacio; exige diversas fuentes
Verificación	Regímenes sectoriales con límites	Evaluaciones específicas y acuerdos parciales
Coste de acceso	Elevado en sistemas estratégicos	Variable; alto en capacidades avanzadas
Prueba pública	Acredita prestaciones delimitadas	Depende de versión y condiciones de ensayo

Fuente: elaboración propia a partir de las obras citadas. IA: inteligencia artificial. Comparación conceptual, no clasificación exhaustiva.

¹⁸ GARCÍA RETUERTA, Oscar y GARCÍA RETUERTA, David. La cadena de valor de la inteligencia artificial: estrategias de autonomía para España. Documento de Opinión IEEE 03/2026. Disponible en: https://www.defensa.gob.es/ceseden/-/ieeee/cadena_de_valor_ia_2026_dieeee03 (consulta: 03/10/2026).

La comparación exige, además, considerar el tiempo. Una demostración informa sobre una versión del sistema en unas condiciones concretas; pierde valor si cambian los datos, las conexiones o el entorno de empleo. Para interpretar una señal no basta con conocer el resultado de una prueba: importa saber qué aspecto de la capacidad acredita y durante cuánto tiempo sigue siendo representativo. Este carácter dinámico complica los acuerdos basados exclusivamente en inventarios y desplaza parte de la atención hacia los procedimientos de evaluación. Se trata de una diferencia de grado y de objeto, no de una desaparición de toda base material.

Implicaciones: opacidad, atribución y erosión de la disuasión

La opacidad afecta al cálculo disuasorio. La disuasión puede basarse en negar al adversario sus objetivos o en amenazar con imponerle costes. Ambas vías requieren que las capacidades y la voluntad de emplearlas resulten suficientemente creíbles. Si una actualización altera prestaciones difíciles de observar, aumenta el margen para subestimar o sobrestimar al rival. Esta incertidumbre puede favorecer errores de cálculo, aunque también inducir prudencia; su efecto no es mecánico y depende de la situación estratégica¹⁹.

La atribución cibernética plantea un problema relacionado. La intermediación, la reutilización de herramientas y la manipulación de indicios ya dificultaban identificar responsables antes de la expansión de la IA²⁰. La IA puede ampliar algunos riesgos, pero también apoyar el análisis defensivo²¹. Conviene distinguir la atribución técnica, la valoración de inteligencia y la decisión política de hacer pública una acusación. Una atribución incierta puede debilitar la amenaza de represalia, sin anular otras formas de disuasión basadas en resiliencia, defensa y cooperación.

La verificación internacional cambia igualmente de objeto. En sistemas de alerta temprana, apoyo al análisis o mando y control nuclear, los componentes algorítmicos

¹⁹ JERVIS, Robert. Perception and Misperception in International Politics, op. cit.

²⁰ BUCHANAN, Ben. The Hacker and the State: Cyber Attacks and the New Normal of Geopolitics. Harvard University Press, Cambridge (MA), 2020.

²¹ BRUNDAGE, Miles et al. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. 2018, arXiv:1802.07228. Disponible en: <https://arxiv.org/abs/1802.07228> (consulta: 03/10/2026).

pueden afectar a la interpretación de información crítica²². Verificar una instalación física no permite conocer necesariamente cómo se comportará un modelo fuera de las condiciones de prueba. Las evaluaciones requieren delimitar funciones, entornos de empleo y criterios de fiabilidad. El acceso a información sensible, además, plantea restricciones políticas y de seguridad que dificultan trasladar directamente los modelos de inspección tradicionales.

El control humano significativo, denominado habitualmente *meaningful human control*, constituye una respuesta relevante, pero debe traducirse en capacidad efectiva de comprensión e intervención. La presencia formal de un operador resulta insuficiente si carece de tiempo, información o autoridad para modificar una decisión. El diseño del sistema debe permitir reconocer límites, gestionar fallos y conservar responsabilidades claras²³. La supervisión humana contribuye a reducir riesgos operativos; no sustituye por sí sola la verificación externa de compromisos internacionales.

Las asimetrías en la cadena de valor añaden una dimensión estructural. El acceso a semiconductores, infraestructura, datos y talento condiciona la autonomía de los Estados, mientras la competencia entre grandes potencias se apoya en indicadores incompletos²⁴. La difusión de ciertas herramientas reduce barreras de entrada, pero desarrollar y sostener capacidades avanzadas puede exigir recursos elevados. De ahí que resulte inadecuado caracterizar toda la IA como una tecnología barata o asumir que el acceso a un modelo equivale a una capacidad militar integrada.

En el espacio informativo, el contenido sintético puede erosionar la confianza en mensajes, pruebas y compromisos públicos. La señalización exige distinguir entre capacidad anunciada, capacidad demostrada y capacidad disponible bajo condiciones operativas. Una respuesta razonable combina autenticación de fuentes, contrastación independiente y comunicación de incertidumbres. El objetivo no es eliminar toda opacidad, algo difícilmente compatible con la seguridad nacional, sino evitar que la ambigüedad haga imposible reconocer límites y compromisos durante una crisis.

²² RODRÍGUEZ LA BARRERA, Luciana Alejandra. Armas nucleares e inteligencia artificial: un equilibrio entre la automatización y el factor humano. Documento de Opinión IEEE 10/2026. Disponible en: https://www.defensa.gob.es/ceseden/-/ieeee/armas_nucleares_2026_dieeee10 (consulta: 03/10/2026).

²³ BOULANIN, Vincent y VERBRUGGEN, Maaïke. Mapping the Development of Autonomy in Weapon Systems, op. cit. SCHARRE, Paul. Army of None, op. cit.

²⁴ SCHARRE, Paul. Four Battlegrounds: Power in the Age of Artificial Intelligence. W. W. Norton, Nueva York, 2023.

Conviene separar la incertidumbre sobre el adversario de la incertidumbre sobre el propio sistema. Es posible que una organización desconozca cómo responderá una herramienta fuera de las condiciones habituales, incluso disponiendo de acceso interno a ella. Si comunica una capacidad a partir de pruebas limitadas, puede transmitir una confianza que tampoco está justificada para sus responsables. La señalización exige, por ello, evaluar tanto la credibilidad externa como la solidez del conocimiento propio. El riesgo de cálculo equivocado no procede únicamente de la ocultación deliberada: también puede surgir de expectativas excesivas, resultados poco representativos o interpretaciones incompatibles de una misma demostración. Esta distinción permite formular el problema sin asumir que toda opacidad responde a una estrategia de engaño.

La adaptación doctrinal debe considerar asimismo las señales defensivas. Demostrar capacidad para mantener servicios esenciales, recuperar funciones o conservar el control de una decisión puede reforzar la disuasión sin revelar detalles de capacidades ofensivas. No se trata de convertir esas demostraciones en garantías universales, sino de hacer explícito qué compromiso respaldan y qué incertidumbres permanecen. En situaciones de crisis, las declaraciones prudentes sobre límites resultan a menudo más útiles que las promesas de autonomía o superioridad general. La credibilidad depende de la coherencia entre el mensaje, la evidencia disponible y la conducta posterior. La IA introduce nuevas dificultades en esa relación, pero no elimina la función del juicio político ni la necesidad de canales de comunicación estables.

El espacio regulatorio y la posición europea

El Reglamento (UE) 2024/1689 establece obligaciones diferenciadas según el riesgo y una aplicación escalonada. Su alcance estratégico tiene un límite esencial: el artículo 2.3 excluye los sistemas en la medida en que se destinen exclusivamente a fines militares, de defensa o de seguridad nacional. Los usos civiles o mixtos requieren analizar el ámbito de aplicación correspondiente²⁵. Por tanto, el Reglamento no constituye un régimen general de control del armamento basado en IA.

²⁵ PARLAMENTO EUROPEO Y CONSEJO. Reglamento (UE) 2024/1689, de 13 de junio de 2024. Diario Oficial de la Unión Europea, 12 de julio de 2024, artículo 2.3, considerando 24 y artículo 113. Disponible en: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj?locale=es> (consulta: 03/10/2026).

España cuenta con la Agencia Española de Supervisión de Inteligencia Artificial (AESIA), cuyo estatuto fue aprobado por el Real Decreto 729/2023²⁶. La Estrategia de Inteligencia Artificial 2024 actualiza el esfuerzo iniciado con la Estrategia Nacional de Inteligencia Artificial de 2020 y vincula capacidad tecnológica, difusión de la IA y responsabilidad²⁷. Estas herramientas pueden aportar experiencia de supervisión y evaluación, pero su existencia no garantiza una capacidad de verificación militar ni una aceptación internacional de los estándares europeos.

En el plano aliado, la Estrategia de IA de la OTAN, revisada en julio de 2024, articula seis principios de uso responsable: legalidad, responsabilidad y rendición de cuentas, explicabilidad y trazabilidad, fiabilidad, gobernabilidad y mitigación de sesgos²⁸. Para España, el reto es conectar interoperabilidad militar, base industrial y prácticas de evaluación, respetando la diferencia entre obligaciones jurídicas civiles y compromisos doctrinales. La gobernabilidad incluye la posibilidad de intervenir ante comportamientos no deseados; no debe reducirse a una declaración genérica de buena gestión.

De esa distinción se desprende una agenda de cooperación que conviene someter a examen. Compartir criterios de evaluación o comunicar incidentes relevantes puede resultar menos intrusivo que exigir acceso completo a modelos, datos y sistemas de mando. Sin embargo, ninguna de estas medidas ofrece garantías automáticas: los participantes deben acordar qué información permite contrastar un compromiso y cómo evitar que su intercambio exponga vulnerabilidades. La reciprocidad, la protección de información sensible y la posibilidad de revisar los mecanismos cuando cambie la tecnología son condiciones de viabilidad. La utilidad estratégica debe juzgarse por su contribución a decisiones más informadas, especialmente durante episodios de tensión.

²⁶ GOBIERNO DE ESPAÑA. Real Decreto 729/2023, de 22 de agosto, por el que se aprueba el Estatuto de la AESIA. BOE núm. 210, de 2 de septiembre de 2023. Disponible en: <https://www.boe.es/eli/es/rd/2023/08/22/729> (consulta: 03/10/2026).

²⁷ GOBIERNO DE ESPAÑA. IA: ¿qué es la Estrategia de Inteligencia Artificial 2024? La Moncloa, 5 de junio de 2024. Disponible en: <https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/transformacion-digital-y-funcion-publica/Paginas/2024/ia-inteligencia-artificial-estrategia-espana.aspx> (consulta: 03/10/2026).

²⁸ OTAN. Summary of NATO's revised Artificial Intelligence (AI) strategy. 10 de julio de 2024. Disponible en: <https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2024/07/10/summary-of-natos-revised-artificial-intelligence-ai-strategy> (consulta: 03/10/2026).

Conclusiones

La señalización estratégica contribuye a la estabilidad cuando permite interpretar capacidades e intenciones con suficiente credibilidad. La defensa frente a amenazas externas forma parte de esa lógica, junto con la posibilidad de imponer costes. La experiencia nuclear y convencional ofrece instrumentos útiles, pero también enseña que la disuasión siempre ha operado con información incompleta y con riesgos de percepción equivocada.

La IA modifica algunos fundamentos de ese intercambio porque debilita la correspondencia entre infraestructura visible y capacidad efectiva. El carácter dual, la actualización del software y la dependencia del contexto de empleo complican la inferencia externa. No desaparecen las huellas físicas ni todas las posibilidades de verificación; cambia la combinación de evidencias necesaria para sostener una evaluación razonable.

La respuesta europea debe reconocer los límites de la regulación civil y desarrollar su articulación con los marcos de defensa. Los debates del Grupo de Expertos Gubernamentales sobre sistemas de armas autónomas letales, en el marco de la Convención sobre Ciertas Armas Convencionales, ofrecen un foro relevante, aunque su objeto no abarca todas las aplicaciones militares de IA²⁹. Trasladar principios a compromisos verificables requiere acuerdos sobre el alcance, las evidencias admisibles y la protección de información sensible.

Como orientación doctrinal, se propone avanzar hacia una transparencia gradual sobre funciones y límites, evaluaciones acordadas y mecanismos de comunicación de incidentes. Estas medidas deben examinarse según su viabilidad técnica, reciprocidad y utilidad en crisis. España puede contribuir mediante capacidades de evaluación, cooperación aliada y participación multilateral. El objetivo es reconstruir condiciones de

²⁹ NACIONES UNIDAS. Convention on Certain Conventional Weapons: Group of Governmental Experts on Lethal Autonomous Weapons Systems (2026). Disponible en: <https://meetings.unoda.org/meeting/79329> (consulta: 03/10/2026).

confianza suficientes para que la innovación tecnológica no amplíe, sin contrapesos, el riesgo de conflicto por error de cálculo.

*Pau Muñoz Pairet **

Doctor en Inteligencia Artificial y en Estudios Estratégicos
Profesor de Relaciones Internacionales, Universidad Pontificia Comillas