



*Solemne acto
apertura curso académico
2026 /27*

Madrid, 13 de octubre de 2026

Doctor D. Rafael Pastor Vargas
Director de la ETSI de Informática en la UNED

El campo de batalla sin trincheras: IA, desinformación y poder

Introducción

Empecemos por una idea incómoda que sirve como punto de partida: la inteligencia artificial que recomienda series de streaming o evalúa la peligrosidad de los correos es, en lo esencial, la misma tecnología que guía un dron o fabrica un vídeo falso de un presidente. La IA es lo que la literatura llama una tecnología de doble uso (Brundage et al., 2018), y eso la convierte en un instrumento de poder militar, político e informativo.

Vamos a repasar cómo se emplea la IA en el campo de batalla, de las trincheras al mundo digital. Primero, un breve resumen de la IA como arma que automatiza la guerra. Después, y ya sin trincheras, el núcleo de esta lección: la influencia y la guerra híbrida, con Ceuta como caso cercano, y la desinformación con DeepFakes, algoritmos y redes coordinadas. Finalmente, la pregunta de fondo: quién mueve los hilos.

***Pregunta para la reflexión:** antes de empezar, ¿cuántas personas del público creen haber visto esta semana algún contenido generado por IA sin saberlo?*

Marco conceptual

Antes de entrar en los casos conviene fijar algunos conceptos, porque en este terreno las palabras se usan con bastante alegría. "IA", "autonomía" o "desinformación" significan cosas distintas según quién las pronuncie, y esa ambigüedad no siempre es inocente.

Casi todo lo que se analiza aquí, se basa en el aprendizaje automático (Machine Learning): sistemas que no se programan regla a regla, sino que aprenden patrones a partir de "enormes" cantidades de datos (Russell y Norvig, 2021). De aquí se deduce una consecuencia importante: una IA es tan buena, o tan sesgada, como los datos con los que se entrena. Y lo que la distingue de revoluciones militares anteriores es que no solo multiplica la fuerza, sino que empieza a sustituir el juicio humano (Kissinger et al., 2021).

La autonomía tampoco es un interruptor de encendido y apagado, sino una escala. La clasificación más extendida distingue tres niveles según la posición del humano respecto al "bucle" de decisión (Scharre, 2018): **dentro** del bucle, cuando el sistema propone y la persona decide cada ataque; **sobre** el bucle, cuando opera solo pero alguien

supervisa y puede interrumpirlo, como en la defensa antimisiles de un buque; y **fuera** del bucle, cuando selecciona y ataca sin intervención. El debate ético y jurídico se concentra en el salto del segundo al tercer nivel, y la frontera entre ambos es cada vez más difusa.

1. La IA como arma en las trincheras: una panorámica

Al pensar en "robots asesinos", la imagen que suele venir a la cabeza es Terminator. La realidad es menos cinematográfica y más inquietante: la IA está reduciendo el papel humano en las tres fases clásicas del combate, que Scharre (2018) resume como ver, decidir y actuar. Horowitz (2019) añade un matiz clave: la IA acelera tanto el combate que los humanos pueden quedar fuera del circuito no por decisión ética, sino por pura velocidad.

1.1 Drones y sistemas autónomos

Los drones son el laboratorio de automatización perfecto porque son baratos y perderlos no cuesta vidas propias. La guerra entre Ucrania y Rusia es, desde 2022, la primera "guerra de drones" a gran escala, donde aparatos FPV de unos cientos/miles de euros destruyen tanques de millones; ante las interferencias electrónicas, ambos bandos han añadido guiado terminal por visión artificial, que mantiene el rumbo, aunque se corte la señal. Rusia emplea en masa los vehículos no tripulados Shahed o Geran-2, de origen iraní, contra la red eléctrica ucraniana, con una lógica que no busca la precisión sino saturar las defensas.

Antes hubo dos avisos. En Nagorno-Karabaj, en 2020, los Bayraktar TB2 turcos y los Harop israelíes dieron a Azerbaiyán una ventaja decisiva. Ese mismo año, en Libia, el Panel de Expertos de la ONU (S/2021/229) indicó que los convoyes de Haftar fueron atacados por sistemas como el STM Kargu-2, "programados para atacar objetivos sin requerir conectividad de datos entre el operador y la munición". Es el caso más citado en el debate sobre armas autónomas, aunque el informe no confirma que causara muertes, así que no debe presentarse como el primer homicidio autónomo probado. Tampoco es todo nuevo: los sistemas antimisiles como la Cúpula de Hierro llevan años funcionando casi en automático, porque un humano no reacciona a tiempo.

Conviene además manejar las cifras de Ucrania con cuidado: Royal United Services Institute for Defence and Security Studies estima que los drones causan entre el 60 % y el 70 % del material ruso destruido (Watling y Reynolds, 2025), mientras que el 70-80 % de bajas humanas son atribuidas por los propios mandos ucranianos. Tampoco

conviene exagerar el papel de la IA, porque la mayoría de los drones siguen siendo pilotados por humanos. La tendencia, eso sí, apunta a una autonomía creciente empujada por la guerra electrónica: cuando el enlace se corta, solo un sistema capaz de decidir por sí mismo completa la misión.

1.2 El software que elige el objetivo

Lo más impactante no es el robot que se ve, sino el software que no se ve y decide dónde golpear. En Gaza, desde 2023, investigaciones de +972 Magazine y Local Call (Abraham, 2024) describieron los sistemas "Lavender" y "The Gospel" (Habsora). Según esas fuentes, Lavender llegó a marcar a unos 37.000 palestinos como presuntos militantes, con un margen de error asumido en torno al 10 % y una revisión humana de unos veinte segundos por objetivo; el ejército israelí lo negó y los presentó como herramientas de apoyo a sus analistas. El precedente estadounidense es el Proyecto Maven, de 2017: Google se retiró en 2018 tras la protesta de sus empleados, pero el proyecto continuó con otras empresas, entre ellas Palantir.

1.3 España y el debate regulatorio

España está en proceso de convertirse en un actor relevante en IA militar. El Ministerio de Defensa aprobó en 2023 su Estrategia de Desarrollo, Implantación y Uso de la Inteligencia Artificial, que insiste en el control humano y en el derecho internacional humanitario; en el plano industrial participa con Francia y Alemania en el FCAS/NGWS, con Indra como coordinadora española, y desarrolla el programa nacional SIRTAP. Diplomáticamente se alinea con la mayoría europea que defiende un "control humano significativo", sin apoyar una prohibición total.

Debido a que no existe tratado alguno que prohíba las armas autónomas letales (LAWS), ya Russell (2019) advierte de que las armas autónomas baratas podrían convertirse en armas de destrucción masiva "escalables". Las posiciones se agrupan en tres bloques: la prohibición total, que defienden Stop Killer Robots y numerosos países latinoamericanos y africanos; la regulación basada en el control humano significativo, que reúne a buena parte de la UE y al Comité Internacional de la Cruz Roja; y la de no crear nuevas normas, sostenida por Estados Unidos, Rusia, Israel o la India.

Pregunta para la reflexión: si un sistema de IA comete un error y mata a civiles, ¿quién es el responsable: el programador, el comandante, la empresa o el Estado?

2. La IA para influir: la guerra híbrida (ya las trincheras se difuminan)

No todas las guerras empiezan con tanques. Hoffman (2007) acuñó el término "guerra híbrida" para describir la mezcla de presión militar, económica, migratoria, cibernética e informativa que se queda justo por debajo del umbral de una guerra abierta. Galeotti (2019) prefiere hablar de "guerra política": el objetivo no es conquistar territorio, sino voluntades. Dicho de forma informal, se trata de dañar a un país sin que este pueda afirmar con claridad que ha sido atacado. Y la IA es el complemento perfecto, porque abarata, acelera y oculta.

2.1 Qué añade la IA en la guerra híbrida

La propaganda y la presión indirecta son tan antiguas como la diplomacia. Lo que cambia con la IA no es la naturaleza de estas tácticas, sino su coste, su velocidad y su capacidad de adaptarse a cada audiencia. El análisis de sentimiento detecta en tiempo real qué temas dividen a una sociedad, ya sea la inmigración, la vivienda, la inflación o la identidad, para meter el dedo en la llaga. La microsegmentación lleva mensajes distintos a cada grupo según su perfil, como hizo Cambridge Analytica en 2016. Y la generación automática produce textos, imágenes y vídeos en decenas de idiomas a coste casi cero.

A ello se suman las ciberoperaciones asistidas y, sobre todo, la negación plausible: los bots, los intermediarios y las cuentas falsas complican atribuir la autoría a un Estado. Rid (2020) recuerda que esto no es nuevo, pues las "medidas activas" soviéticas ya lo practicaban, pero la escala sí lo es.

Un apunte sobre el terreno cibernético, porque no todas las armas hacen ruido. Rid (2013) ya advertía de que la "ciberguerra" rara vez es guerra en sentido estricto: se parece más al sabotaje y al espionaje, aunque sus efectos pueden ser devastadores, como demostraron Stuxnet en 2010 o NotPetya en 2017, con pérdidas de más de 10.000 millones de dólares. La novedad es que en noviembre de 2025 Anthropic describió una campaña de ciberespionaje atribuida a un grupo patrocinado por el Estado chino en la que un agente de IA ejecutó entre el 80 % y el 90 % de las tareas tácticas.

2.2 El caso de Ceuta (mayo de 2021)

Entre el 17 y el 19 de mayo de 2021 entraron en Ceuta unas 10.000 personas, muchas de ellas menores, a nado o bordeando los espigones, mientras el control fronterizo marroquí se relajaba de forma visible. El detonante puede decirse que fue, entre otros, la acogida en

un hospital español del líder del Frente Polisario, Brahim Ghali, que Marruecos consideró una afrenta. El Gobierno español y el Parlamento Europeo (2021) lo interpretaron como presión política que usaba la migración como instrumento; Marruecos rechazó esa lectura y lo enmarcó como una crisis de confianza bilateral. El conflicto parece que se cerraba en 2022, cuando España apoyó el plan de autonomía marroquí para el Sáhara Occidental.

Greenhill (2010) ya había teorizado este fenómeno con el concepto de "armas de migración masiva": Estados que provocan o toleran flujos migratorios para arrancar concesiones. Ceuta encaja en el patrón por tres razones: la migración instrumentalizada, con personas convertidas en palanca de presión sin usar armas; la batalla informativa paralela, con vídeos, rumores y cuentas coordinadas que alimentaban tanto el pánico como el discurso xenófobo; y el objetivo político de forzar un cambio de posición diplomática que, visto el resultado, cabe preguntarse si funcionó.

Ojo con exagerar: el papel concreto de la IA en Ceuta es limitado y difícil de probar. Lo relevante es que con las herramientas actuales una operación similar sería hoy mucho más potente. La Estrategia de Seguridad Nacional (Gobierno de España, 2021) ya situó las amenazas híbridas entre los riesgos prioritarios.

Pregunta para la reflexión: ¿dónde está la línea entre una decisión política legítima de un país sobre su frontera y un acto de guerra híbrida?

2.3 Otros ejemplos de guerra híbrida

Ceuta no es un caso aislado. En la última década se repite un mismo patrón en escenarios muy distintos: presión por debajo del umbral bélico, difícil atribución y una campaña informativa que acompaña a la acción física.

En Crimea, en 2014, los "hombrecillos verdes", tropas sin distintivos, actuaron acompañados de redes de cuentas y medios estatales que negaban la intervención. En 2016, la injerencia en las elecciones estadounidenses se apoyó en los miles de perfiles falsos de la Internet Research Agency rusa. Durante el referéndum catalán de 2017, informes europeos y del Gobierno español señalaron cuentas vinculadas a RT y Sputnik amplificando la polarización, aunque su alcance sigue discutido. En 2021, Bielorrusia instrumentalizó la migración hacia Polonia promocionando rutas "fáciles" hacia la UE. Y la campaña "Doppelganger" (2022-2024) suplantó a medios europeos con clones de Le Monde o Der Spiegel, en parte generados con IA. La

UE respondió con el East StratCom Task Force (2015) y EUvsDisinfo, y la OTAN reconoce desde 2016 el ciberespacio como dominio de operaciones.

Pregunta para la reflexión: *Bielorrusia-Polonia y Ceuta se parecen mucho, ¿por qué la reacción mediática fue tan distinta?*

2.4 La guerra cognitiva: la mente como campo de batalla

Si la guerra híbrida busca debilitar a un país sin declararle la guerra, la guerra cognitiva va un paso más allá: su objetivo es la forma en que la población piensa, siente y decide. La OTAN empezó a estudiar el concepto a partir de 2020, y un informe de su centro de innovación (Du Cluzel, 2021) llegó a describir el cerebro humano como "el campo de batalla del siglo XXI".

La idea puede sonar exagerada, pero tiene una lógica sencilla: una sociedad dividida, desconfiada y saturada de información contradictoria es más fácil de manipular y más difícil de movilizar ante una amenaza real. No hace falta convencer a nadie de una mentira en particular; basta con que deje de creer en todo.

La IA encaja aquí de tres maneras. Permite un mejor conocimiento del adversario, porque el análisis masivo de redes revela qué miedos e identidades pesan más en cada grupo. Facilita la personalización, adaptando los mensajes a cada perfil psicológico. Y aporta persistencia, ya que los sistemas automatizados sostienen campañas durante años sin picos que llamen la atención. China lleva desde 2003 incorporando a su doctrina las "tres guerras": la de la opinión pública, la psicológica y la jurídica, con Taiwán como laboratorio. Conviene, eso sí, una cautela: algunos académicos critican que el concepto es tan amplio que corre el riesgo de convertir cualquier discrepancia política en una "amenaza".

2.5 Las infraestructuras críticas como blanco híbrido

La otra gran vía de presión híbrida no se dirige a la mente, sino a las infraestructuras críticas de las que depende la vida diaria: electricidad, agua, telecomunicaciones, transporte o cables submarinos. La ventaja para el atacante es que un daño pequeño y difícil de atribuir puede generar un gran impacto social y político.

El caso más conocido es el de Nord Stream: en septiembre de 2022, varias explosiones inutilizaron los gasoductos que unían Rusia y Alemania bajo el mar Báltico, y la autoría sigue siendo objeto de controversia, lo que ilustra a la perfección el *problema de la atribución*.

Entre 2023 y 2025, varios cables del Báltico resultaron dañados por anclas arrastradas por buques vinculados a China o a la "flota fantasma" rusa; Finlandia incautó uno de ellos, el Eagle S, en diciembre de 2024, aunque un tribunal de Helsinki desestimó después la causa penal por falta de jurisdicción.

El episodio más reciente llegó desde el aire. En la noche del 9 al 10 de septiembre de 2025 se registraron cerca de una veintena de violaciones del espacio aéreo polaco por drones rusos y Polonia derribó varios, la primera vez que lo hacía un aliado de la OTAN, lo que llevó a activar el artículo 4 del Tratado. La IA interviene en estos ataques de dos formas: como herramienta para localizar vulnerabilidades en sistemas de control industrial y como motor de las campañas de desinformación que acompañan a cada incidente, amplificando el miedo y embarrando la atribución.

***Pregunta para la reflexión:** ante un apagón o un corte de internet, ¿cuánto tiempo tardaría la sociedad en empezar a creer rumores sobre su causa?*

3. La IA y la desinformación: el poder de la manipulación

Wardle y Derakhshan (2017) distinguen tres elementos que suelen meterse en el mismo saco: la información errónea (falsa, pero sin intención), la desinformación (falsa y deliberada) y la malinformación (verdadera, pero usada para hacer daño y habitualmente sacada de contexto). La IA generativa potencia las tres, y sobre todo abarata la segunda: lo que antes exigía un estudio de producción ahora cabe en un portátil.

Pero el efecto más preocupante es otro. Chesney y Citron (2019) lo llaman el "dividendo del mentiroso": cuando todo puede ser falso, cualquiera puede decir que un vídeo real es un Deepfake (o bulo, elijan su preferencia). Es decir, la IA no solo hace creer en las mentiras, también hace dudar de las verdades.

3.1 Deepfakes

Son vídeos, audios o imágenes sintéticos que imitan a personas reales. En marzo de 2022, un vídeo falso mostraba a Zelenski pidiendo a sus tropas que se rindieran; era burdo y se desmintió en horas, pero fue el primer Deepfake político en plena guerra. En septiembre de 2023, dos días antes de las elecciones eslovacas, circuló un audio falso de un líder opositor hablando de comprar votos, en pleno silencio electoral, justo cuando era más difícil desmentirlo. En enero de 2024, llamadas robóticas con la voz falsa de Biden pedían no votar en las primarias de New Hampshire: la FCC (Comisión Federal de

Comunicaciones) multó con seis millones de dólares al consultor responsable, y la operadora que cursó las llamadas pagó otro millón en un acuerdo.

No todo ocurre tan lejos. En Almendralejo, en 2023, se difundieron desnudos falsos generados con IA de más de veinte chicas menores a partir de fotos reales, y en julio de 2024 el Juzgado de Menores de Badajoz condenó a quince menores a un año de libertad vigilada. Tampoco todo es política: en 2024, un empleado de la ingeniería Arup en Hong Kong transfirió unos 25,6 millones de dólares tras una videollamada en la que todos los "directivos" eran Deepfakes.

***Pregunta para la reflexión:** si mañana circulara un audio de un político español confesando un delito, ¿qué debería hacer cualquier persona antes de compartirlo?*

3.2 Manipulación algorítmica

Los algoritmos de recomendación deciden automáticamente qué ve o escucha cada persona en función de los temas en los que invierte más tiempo. Priorizan lo que engancha, y lo que más engancha suele ser lo que indigna o produce odio. Usan nuestra parte más emocional. Vosoughi, Roy y Aral (2018) mostraron en un estudio con millones de tuits que las noticias falsas se difundían más lejos y más rápido que las verdaderas, sobre todo por su carácter novedoso y emocional.

El caso más grave es el de Myanmar: en 2017, la ONU señaló que Facebook tuvo un papel relevante en la difusión del odio contra los Rohinyá. En el terreno teórico, Pariser (2011) popularizó la idea de las **burbujas de filtro**, según la cual los algoritmos encierran a cada usuario en su propia visión del mundo, aunque estudios posteriores con datos de Meta (Guess et al., 2023) indican que cambiar el algoritmo no altera tanto las actitudes políticas como se pensaba.

El debate más reciente afecta a dos plataformas. En X, antes Twitter, los cambios de algoritmo tras su compra en 2022 alteraron qué cuentas ganan visibilidad, algo que la Comisión Europea investiga bajo la Ley de Servicios Digitales. Y en diciembre de 2024 el Tribunal Constitucional rumano anuló la primera vuelta de las presidenciales que había ganado Călin Georgescu, a partir de documentos desclasificados sobre una campaña coordinada en TikTok; los comicios se repitieron en mayo de 2025.

3.3 Redes coordinadas y Bots

Estas redes se componen de grupos de cuentas (automatizadas, falsas o reales pagadas) que actúan juntas para fingir un apoyo o un

rechazo masivo. Woolley y Howard (2018) lo llaman "propaganda computacional", y Ferrara et al. (2016) ya advertían de lo difícil que es distinguir a un bot de una persona. Con la IA actual, cada cuenta escribe con estilo propio y sin faltas, así que la tarea es aún más complicada.

El ejemplo de mayor tamaño conocido es Spamouflage, también llamada Dragonbridge: una red china de miles de cuentas con mensajes favorables a Pekín. En agosto de 2023, Meta retiró unas 7.700 cuentas y 954 páginas vinculadas, y la describió como la mayor operación de influencia encubierta multiplataforma conocida hasta entonces. OpenAI informó entre 2024 y 2025 de actores de Rusia, China, Irán e Israel que usaban sus modelos para redactar comentarios y generar artículos, aunque con alcance limitado.

Un caso particular es la red rusa Pravda, cientos de webs que publican propaganda en masa y cuyo objetivo, según los investigadores, es también "contaminar" los datos con los que se entrenan los Chatbots. La desinformación, además, no es solo cosa de potencias extranjeras: en Filipinas, la India o Nigeria hay empresas que venden campañas de influencia a los propios partidos políticos.

3.4 Los Chatbots: nuevo vector y nuevo blanco

Hasta hace poco, la desinformación circulaba por redes sociales y webs. Desde la llegada de ChatGPT a finales de 2022, cada vez más personas consultan directamente a un asistente de IA para informarse. Esto abre un frente nuevo, porque el Chatbot se convierte a la vez en posible fuente de errores y en objetivo de manipulación.

Los riesgos principales son tres. El primero son las alucinaciones: los modelos pueden inventar datos, citas o referencias con total seguridad aparente. Es una limitación técnica y comercial (se programan para dar siempre una respuesta, exista o no), pero el efecto sobre quien no verifica es el mismo. El segundo es la contaminación de datos: si alguien inunda internet de propaganda, parte de ella acaba en el entrenamiento o en las búsquedas de los chatbots, y una auditoría de NewsGuard (2025) concluyó que los diez principales asistentes repetían narrativas de la red rusa Pravda en el 33 % de las pruebas, estrategia bautizada como "LLM grooming".

El tercer riesgo es la persuasión: varios estudios apuntan a que los modelos de IA pueden ser tan persuasivos como los humanos, o más. Pero funciona en ambas direcciones, y Costello, Pennycook y Rand (2024) mostraron en *Science* que conversar con un modelo de IA reducía de forma duradera la creencia en teorías conspirativas en

torno a un 20 %. La misma herramienta que puede engañar puede también desmentir.

***Pregunta para la reflexión:** ¿qué nivel de confianza merece una respuesta de un chatbot frente a la de un periódico o la de un experto?*

3.5 La desinformación en España: dos casos recientes

Dos crisis recientes muestran cómo funciona la desinformación en España cuando la población está asustada y busca respuestas rápidas.

La DANA de Valencia (octubre de 2024). Las inundaciones del 29 de octubre causaron 229 muertos en la Comunidad Valenciana, y más de 230 sumando Castilla-La Mancha y Andalucía, y fueron seguidas de una oleada de bulos. Uno de los más difundidos aseguraba que el aparcamiento del centro comercial Bonaire estaba lleno de cadáveres ocultos por las autoridades; cuando la UME y la Guardia Civil lo inspeccionaron, no encontraron ninguna víctima. Circularon también imágenes generadas con IA y cifras inventadas, y Maldita.es y Newtral documentaron cientos de contenidos falsos en pocas semanas. Cabe aquí una pregunta adicional: ¿quién verifica al verificador?

El apagón del 28 de abril de 2025. Un corte eléctrico dejó sin luz a casi toda la península ibérica durante horas. En pleno vacío informativo se multiplicaron las teorías: ciberataques, sabotaje, un fenómeno atmosférico extraño o una conspiración contra las renovables. El informe del comité de análisis del Gobierno, publicado el 17 de junio de 2025, concluyó que se trató de un fenómeno de sobretensiones en cascada por una capacidad insuficiente de control de tensión, y descartó el ciberataque. Para entonces, muchas de aquellas versiones ya estaban asentadas.

Ambos casos comparten un patrón conocido: la desinformación prospera en los huecos que dejan la incertidumbre y la lentitud de la información oficial. La IA no inventó esa forma de actuar, pero la acelera, porque fabricar una imagen creíble o un audio alarmista cuesta hoy segundos.

***Pregunta para la reflexión:** en una emergencia, ¿quién debería informar primero y más rápido: el Gobierno, los medios o los propios ciudadanos desde el lugar?*

3.6 Cómo se detecta y se combate

Ninguna respuesta basta por sí sola, pero combinadas ayudan. En el plano regulatorio destacan la Ley de Servicios Digitales y el

Reglamento de IA europeo, que obliga a etiquetar los Deepfakes. En el tecnológico se trabaja en marcas de agua y metadatos de origen (estándar C2PA) y en detectores de contenido sintético. La verificación profesional corre a cargo de Maldita.es, Newtral y EFE Verifica en España, o el EDMO a escala europea. Y en el terreno educativo está la alfabetización mediática desde la escuela, donde Finlandia es el ejemplo más citado.

Pregunta para la reflexión: *¿es preferible que las plataformas borren el contenido falso o que lo etiqueten y dejen decidir al usuario? ¿Quién debería decidir qué es falso?*

4. Quién dirige estos usos: gobiernos, grandes tecnológicas e industria de defensa

Queda la pregunta que está detrás de todo lo anterior: ¿quién decide para qué se usa la IA? La respuesta corta es que muy pocos. Crawford (2021) insiste en que la IA no es algo etéreo que vive "en la nube", sino una industria material que depende de minerales, chips, energía, datos y trabajo humano. Y quien controla esos recursos controla la tecnología: los chips dependen de Nvidia, de TSMC en Taiwán y de ASML en los Países Bajos; la nube, de Amazon, Microsoft y Google, que alojan tanto a los laboratorios de IA como los contratos militares; y la energía es cada vez más decisiva, como ilustra el acuerdo de Microsoft de 2024 sobre la central de Three Mile Island.

Los Estados juegan tres papeles a la vez: clientes, reguladores y competidores. Lee (2018) describió la relación entre EE. UU. y China como una carrera entre dos "superpotencias de la IA". Estados Unidos restringe desde 2022 la exportación de chips avanzados a China, mientras esta aplica la "fusión civil-militar", por la que las empresas privadas deben colaborar con el Estado y el Ejército. La UE apuesta por regular mediante el Reglamento de IA, en vigor desde 2024, aunque excluye los usos militares mientras sus planes de rearme incluyen tecnologías autónomas. ¿Contradicción o puro realismo?

Hace diez años las tecnológicas presumían de "no hacer el mal". Hoy la frontera entre empresa de software y contratista militar es cada vez más borrosa: Palantir nació con financiación de la CIA, y entre 2024 y 2025 los propios laboratorios de IA relajaron sus prohibiciones sobre usos militares, justificándolo como apoyo a las democracias frente a rivales autoritarios. Ese poder tampoco se ejerce solo hacia fuera: Feldstein (2019) identificó al menos 75 países que ya usaban vigilancia basada en IA, y en España el caso Pegasus reveló en 2022

un espionaje con al menos 65 personas afectadas, incluidos el presidente del Gobierno y varios ministros.

Existen dos corrientes diferenciadas sobre todo ello. La visión crítica (Zuboff, 2019; Crawford, 2021) denuncia que unas pocas empresas y Estados usen la IA para extraer datos, controlar la información y reforzar su poder sin rendir cuentas. La visión estricta de seguridad replica que, si las democracias renuncian a la IA militar, regalarán la ventaja a los regímenes autoritarios. Entre ambas está la visión reguladora, que admite los usos defensivos pero exige supervisión humana, transparencia y límites internacionales, como se hizo con las armas químicas.

Pregunta para la reflexión: ¿qué resulta más preocupante: que la IA la controlen unos pocos gobiernos o unas pocas empresas? ¿Hay alguna diferencia real entre ambas cosas?

5. Escenarios de futuro y conclusiones

Nadie sabe con certeza hacia dónde irá la IA, pero sí es posible dibujar escenarios. El primero es el de una carrera sin frenos, en el que Estados Unidos y China compiten sin límites y las armas autónomas proliferan hacia actores no estatales. El segundo es el de los bloques tecnológicos, con un mundo dividido en ecosistemas separados. El tercero es el de una gobernanza pactada, con acuerdos internacionales y estándares de transparencia, que anuncian las resoluciones de la ONU y el Reglamento europeo. Y el cuarto, quizá el más probable y el menos espectacular, es el de una erosión silenciosa: sin grandes guerras de robots, pero con una desconfianza generalizada hacia la información y las instituciones. Lo más probable es una combinación de los cuatro.

Existen esfuerzos reales de gobernanza, aunque modestos: las cumbres de Bletchley Park (2023), Seúl (2024) y París (2025) reunieron a gobiernos y empresas, si bien la de París dejó ver las divisiones, con Estados Unidos y el Reino Unido fuera de la declaración final que China sí suscribió. En cuanto a las propuestas, a los Estados les correspondería negociar un tratado sobre armas autónomas con control humano significativo; a la Unión Europea, reforzar su soberanía tecnológica; a las empresas, auditorías externas y etiquetado del contenido sintético; y al sistema educativo, la alfabetización mediática y algorítmica desde la escuela.

Si hubiera que resumirlo en una frase: la IA está transformando el conflicto en tres niveles que se retroalimentan. Automatiza el combate,

abaratada la presión híbrida y multiplica la desinformación. Y detrás de los tres hay una misma estructura: pocos Estados y pocas empresas que controlan la tecnología y deciden sus límites. Los casos analizados (Ucrania, Gaza, Ceuta, Eslovaquia, Rumanía) demuestran que no es ciencia ficción, sino algo que ya está pasando. La gran pregunta es si la regulación, la transparencia de las plataformas y la alfabetización ciudadana podrán ir al ritmo de la tecnología. Hasta ahora, la tecnología va ganando la carrera.

Pregunta final para la reflexión: después de todo esto, ¿qué puede hacer la ciudadanía para no convertirse en una pieza más de este tablero?

Referencias

Los datos se han verificado contra fuentes primarias. Las cifras aportadas por los propios beligerantes no son verificables de forma independiente, por lo que se citan con su fuente.

- Abraham, Y. (2024). 'Lavender': The AI machine directing Israel's bombing spree in Gaza. *+972 Magazine*.
- Brundage, M., Avin, S., Clark, J., et al. (2018). *The malicious use of artificial intelligence*. Future of Humanity Institute et al.
- Chesney, R., y Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753-1820.
- Costello, T. H., Pennycook, G., y Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, 385(6714), eadq1814.
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Du Cluzel, F. (2021). *Cognitive warfare*. NATO Allied Command Transformation Innovation Hub.
- Feldstein, S. (2019). *The global expansion of AI surveillance*. Carnegie Endowment for International Peace.
- Ferrara, E., Varol, O., Davis, C., Menczer, F., y Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96-104.
- Galeotti, M. (2019). *Russian political war: Moving beyond the hybrid*. Routledge.
- Gobierno de España. (2021). *Estrategia de Seguridad Nacional 2021*. Departamento de Seguridad Nacional.
- Greenhill, K. M. (2010). *Weapons of mass migration*. Cornell University Press.

- Guess, A. M., Malhotra, N., Pan, J., et al. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*, 381(6656), 398-404.
- Hoffman, F. G. (2007). *Conflict in the 21st century: The rise of hybrid wars*. Potomac Institute for Policy Studies.
- Horowitz, M. C. (2019). When speed kills: Lethal autonomous weapon systems, deterrence and stability. *Journal of Strategic Studies*, 42(6), 764-788.
- Kissinger, H. A., Schmidt, E., y Huttenlocher, D. (2021). *The age of AI: And our human future*. Little, Brown and Company.
- Lee, K.-F. (2018). *AI superpowers: China, Silicon Valley, and the new world order*. Houghton Mifflin Harcourt.
- Naciones Unidas, Panel de Expertos sobre Libia. (2021). *Informe final* (S/2021/229, párr. 63). Consejo de Seguridad.
- NewsGuard. (2025, marzo). *A well-funded Moscow-based global 'news' network has infected Western AI tools with Russian propaganda*.
- Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin Press.
- Parlamento Europeo. (2021). *Resolución, de 10 de junio de 2021, sobre el uso de menores por las autoridades marroquíes en la crisis migratoria de Ceuta* (2021/2747(RSP)).
- Rid, T. (2013). *Cyber war will not take place*. Hurst.
- Rid, T. (2020). *Active measures: The secret history of disinformation and political warfare*. Farrar, Straus and Giroux.
- Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking.
- Russell, S., y Norvig, P. (2021). *Artificial intelligence: A modern approach* (4.^a ed.). Pearson.
- Scharre, P. (2018). *Army of none: Autonomous weapons and the future of war*. W. W. Norton.
- Vosoughi, S., Roy, D., y Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.
- Watling, J., y Reynolds, N. (2025). *Tactical developments during the third year of the Russo-Ukrainian war*. RUSI.
- Wardle, C., y Derakhshan, H. (2017). *Information disorder*. Consejo de Europa.
- Woolley, S. C., y Howard, P. N. (Eds.). (2018). *Computational propaganda*. Oxford Univ. Press.
- Zuboff, S. (2019). *The age of surveillance capitalism*. PublicAffairs. (Trad. esp.: Paidós, 2020).

