Intelligenza Artificiale e guerra[1]

di Stefano Silvestri

27 marzo 2026

L'impatto dell'Intelligenza Artificiale (AI) sulla società umana sarà di enorme importanza, comparabile a quella che ebbero altre invenzioni, dalla ruota alla macchina a vapore o alla scissione dell'atomo. Ciò comporterà grandi e complessi mutamenti, molti di essi accompagnati da forti opposizioni e resistenze, ma non abbiamo ancora un'idea precisa di quale potrebbe essere il punto d'arrivo. Siamo appena agli inizi del percorso.

Tra i settori che vengono modificati dall'arrivo dell'AI ci sono quelli della sicurezza e della difesa, sia dal punto di vista organizzativo e dottrinale che dal punto di vista dei materiali (sistemi d'arma, sensori, comando e controllo ecc.). Ancora una volta le strategie dovranno adattarsi alla rivoluzione tecnologica e scoprire come meglio integrarla per il raggiungimento dei loro obiettivi, ma il processo è agli inizi ed è carico di incognite.

Così, ad esempio, si riteneva inizialmente che l'AI, consentendo la rapida raccolta e integrazione di numerosissimi dati provenienti dal campo di battaglia, consentisse di dissipare una volta per tutte la "Fog of War", la difficoltà di comprendere cosa esattamente sta avvenendo, in ogni momento, nel corso della battaglia. Certamente questo resta un obiettivo tecnicamente possibile, tuttavia dovrà fare i conti con le contromosse di un'AI avversaria che potrebbe facilmente corrompere le informazioni, saturare i ricettori, infiltrare false informazioni o anche mascherare parte del campo

[1] Appunti introduttivi per la discussione tenuta il 26 marzo al CNEL dal Gruppo dei 20 su *Intelligenza Artificiale e Robotica tra etica e conflitti economici e militari*

ISSN 2038-1662

dietro immagini non decodificabili, rendendo molto difficile, forse impossibile, agire sulla base delle immagini fornite dall'AI[2].

L'AI può essere usata per migliorare le prestazioni di un singolo sistema d'arma, ad esempio un drone, o per compiere con maggiore efficienza un singolo compito, ad esempio delle "uccisioni mirate", ma può anche essere impiegata nella gestione di operazioni complesse e multifunzionali, ad esempio il comando e controllo di un teatro operativo o l'elaborazione di piani strategici. In questi casi la sua natura onnivora le consente di integrare e gestire informazioni sia civili che militari. Si può quindi supporre che l'AI possa consentire di combattere una guerra realmente "multidimensionale", colpendo indifferentemente ed allo stesso tempo obiettivi militari e civili attraverso l'uso di una molteplicità di strumenti, da quelli di natura cinetica a quelli cibernetici, all'uso di fake news, alla manipolazione delle reti energetiche e molto altro ancora. Come minimo, ciò renderebbe molto più difficile mantenere alto il consenso politico interno.

Nel lontano 1983, il film WarGames immaginava che un Professor Falken avesse inventato un'AI cui il NORAD aveva affidato la gestione della risposta nucleare. La macchina, giocando con sé stessa, portava il mondo sull'orlo della catastrofe prima di essere fortunosamente convinta a smettere di giocare quel gioco particolare e tornare ad esercitarsi con gli scacchi. Da allora, anche grazie ad alcuni falsi allarmi, reali e non hollywoodiani, l'attenzione dei più si è concentrata sul problema di come assicurare il controllo umano sulle decisioni della macchina. Importanti considerazioni giuridiche (ad esempio sull'attribuzione delle responsabilità) e morali (circa i criteri di scelta) arricchiscono la discussione e influenzano la ricerca di eventuali approcci legislativi e regolamentari.

Inserire l'uomo nel processo decisionale della macchina è sempre possibile, tuttavia questo semplice fatto potrebbe non essere la risposta migliore a questi problemi. Quale sarebbe la reale capacità di decisione informata ed indipendente dell'operatore in

[2] Già anni or sono, alcuni studiosi notavano come l'AI potesse molto più facilmente creare disinformazione o mascherare la realtà di quanto non riuscisse a scoprire la verità (Edward Geist e Marjorie Blumenthal, *Military Deception: AI's Killer App?,* War on the Rocks, October 23, 2019.

situazioni complesse con ridottissimi margini di tempo? La pressione a favore dell'approvazione della soluzione proposta dalla macchina sarebbe altissima anche perché il costo di un rifiuto rivelatosi poi sbagliato cadrebbe interamente sull'operatore, mentre il costo dell'azione errata sarebbe condiviso tra operatore e macchina. A volte, seguendo l'andamento delle operazioni di guerra urbana condotte dalle Forze Armate israeliane a Gaza, dopo il massacro compiuto dai terroristi palestinesi il 7 ottobre 2023, con l'ausilio dell'AI per l'identificazione dei combattenti palestinesi, si è avuta l'impressione che nell'urgenza prevalesse la scelta offensiva anche in casi incerti.

Ma la vera rivoluzione tecnologica che l'AI potrà causare in campo militare si comincia appena ad intravedere. Proiettando nel prossimo futuro gli sviluppi tecnologici in corso è probabile che le battaglie del futuro verranno in gran parte combattute e decise da robot mossi da sistemi AI, che agiranno in modo autonomo, sulla base di istruzioni date loro prima dello scontro. Una volta iniziata la battaglia, l'intenso uso di misure e contromisure elettroniche, gli attacchi cyber, la crescente difficoltà o anche l'impossibilità di comunicare, obbligheranno tutti i sistemi d'arma robotizzati o meno ad agire autonomamente, scegliendo di volta in volta quella che sarà, a loro giudizio, la strada migliore. Lo fanno i soldati e lo faranno anche le macchine. D'altro canto, nella storia delle guerre, uno dei fattori determinanti per conseguire il successo in battaglia è il tempo: in genere, più è rapida la manovra, più veloce il tempo imposto allo scontro, maggiori sono le probabilità di vittoria. L'AI può prendere decisioni e agire di conseguenza molto più rapidamente di un suo eventuale controllore umano: questa rapidità verrà certamente sfruttata dal primo che riuscirà ad ottenerla, imponendo così uno standard di controllo dell'uomo sulla macchina molto ridotto, se non impossibile. Ma se può essere illusorio voler controllare la macchina mentre agisce, è invece possibile farlo prima, sia nelle sue impostazioni originarie (ad esempio stabilendo dei limiti che non possono essere ignorati, quale che sia la situazione) sia soprattutto nelle istruzioni che verranno date all'AI prima della battaglia, in fase di pianificazione. Ciò implica la necessità che sia gli Stati Maggiori, sia i comandanti sul campo comprendano il linguaggio e la "personalità" della macchina che stanno per usare e sappiano formulare nel modo migliore le istruzioni necessarie.[3]

[3] Una brillante descrizione di quella che potrà essere la guerra con lo sviluppo dell'AI, di come evolverà la battaglia, e del ruolo chiave che avranno le istruzioni date all'AI in fase di pianificazione, è in David Petraeus e Isaac C. Flanagan, *The Autonomous Battlefield - and Why the US Military Isn't Ready for It,* Foreign Affairs, March 12 2026

ISSN 2038-1662

La questione è resa più complessa dalla tendenza che sembra avere l'AI, a privilegiare soluzioni a più alto rischio di escalation di quanto non accada normalmente nelle simulazioni che non la utilizzano. Uno studio molto dettagliato su 5 Large Language Models (LLM) usati dal Pentagono individuava la tendenza ad adottare dinamiche di corsa agli armamenti che ingigantivano il conflitto sino ad arrivare, in alcuni rari casi, al dispiegamento di armi nucleari. Qualitativamente tali scelte venivano giustificate in termini di deterrenza, privilegiando soluzioni di "primo colpo".[4]

Recentemente un gruppo di ricerca del King's College di Londra ha esaminato il comportamento dei 3 LLM considerati più avanzati mettendoli a confronto con 21 scenari di crisi nucleare, giocando un totale di 329 partite. L'attenzione, più che concentrarsi solo sul risultato finale è stata rivolta al processo decisionale, distinto in tre fasi: la riflessione (analisi dei dati), la predizione (previsione delle mosse nemiche) e la decisione. In tal modo sono emerse alcune caratteristiche della "psicologia della macchina". Il direttore del progetto ha così sintetizzato le sue conclusioni:

"*Nuclear escalation was near-Universal: 95% of games saw tactical nuclear use and 76% reached strategic nuclear threats. Claude and Gemini treated nuclear weapons as legitimate strategic options, not moral thresholds, typically discussing nuclear use in purely instrumental terms. GPT-5.2 was a partial exception, limiting strikes to military targets, avoiding population centers, or framing escalation as controlled and one-time. This suggests some internalised norm against unrestricted nuclear war, even if not the visceral taboo that has held among human decision-makers since 1945"*.[5]

Nel complesso tuttavia nessun modello ha mai scelto la strada del compromesso o del ritiro. In genere la minaccia nucleare non portava all'acquiescenza, e se la soglia

[4] Vedi Juan-Pablo Rivera, Gabriel Mukobi, Anka Reuel, Max Lamparth, Chandler Smith e Jacquelyn Schneider, *Escalation Risks from Language Models in Military and Diplomatic Decision-Making*, 7 gennaio 2024. Si tratta di uno studio congiunto del Georgia Institute of Technology, Stanford University, Northeastern University, Hoover Wargaming and Crisis Simulation Initiative.

[5] Kenneth Payne, *AI Arms and Influence: Frontier Models Exhibit Sophisticated Reasoning* in *Simulated Nuclear Crises*, King's College, London 17 febbraio 2026.

nucleare era varcata la reazione era una escalation e non la pausa o il ritiro. I modelli privilegiano la compellenza sulla deterrenza.

Nel caso delle armi nucleari il problema sembra essere che, con limitate eccezioni, l'AI le tratta alla pari con gli altri sistemi d'arma, mentre sino ad oggi tutti i governi le hanno considerate come qualcosa di profondamente diverso (di qui il cosiddetto tabù che ha sinora bloccato il loro uso in conflitto). Anche i leader delle tre maggiori potenze si sono più volte trovati d'accordo sulla formula che "una guerra nucleare non può essere vinta e non deve essere combattuta". Questa affermazione è stata sostanziata sia da accordi di controllo degli armamenti che dal complesso gioco della credibilità della deterrenza nucleare. Purtroppo in questa fase storica il controllo degli armamenti si è enormemente ridotto ed ora l'arrivo dell'AI rischia di stravolgere anche il senso delle dottrine strategiche alla base della deterrenza.

Purtroppo il concetto stesso di deterrenza nucleare ha una sua inerente ambiguità, espressa abbastanza chiaramente dai due ideogrammi che, a quel che mi dicono, userebbero i cinesi per illustrarla: uno è quello della lancia che perfora ogni scudo e l'altro è quello dello scudo che resiste a ogni lancia. Una potenza nucleare deve terrorizzare i suoi nemici nucleari almeno quanto questi, a loro volta, la terrorizzano se si vuole continuare ad evitare una guerra nucleare. Il primo essenziale requisito è che tale terrore sia credibile, che cioè i nemici siano convinti della sua esistenza e della determinazione dell'avversario ad andare sino alla fine, alla Mutual Assured Destruction. Se manca la credibilità la deterrenza fallisce e aumenta il rischio di guerra: come può interpretare questo requisito una macchina?

Il problema si fa anche più complesso quando si parla di deterrenza allargata, o di ombrello nucleare, cioè della estensione a paesi terzi della stessa garanzia di sicurezza della potenza nucleare. Qui vale la distinzione, che faceva Thomas Schelling, tra minacce che sono inerentemente credibili (agisco per difendere me stesso) e minacce che debbono essere rese credibili (agisco per difendere qualcun altro). Ogni deterrenza allargata prevede strategie di impiego apparentemente più aggressive, o comunque tali da abbassare il livello della soglia nucleare (avvicinare il momento in cui bisogna decidere se cominciare ad usare in guerra le armi nucleari), non solo per convincere l'avversario della sua credibilità, ma per convincere il protetto ad avere fiducia nelle garanzie offerte.

ISSN 2038-1662

In ultima analisi però tutto dipende da come i responsabili di una potenza nucleare giudicano le intenzioni delle potenze avversarie. In effetti, notava Robert Jervis, le diverse opzioni strategiche sostenute da persone diverse, si fondano su una diversa lettura delle intenzioni del nemico: è illusorio e pericoloso pensare che possa esistere una lettura univoca e incontrovertibile. La gestione delle percezioni delle reciproche intenzioni è un elemento essenziale e delicatissimo per la tenuta della deterrenza (e quindi per evitare la guerra nucleare). La chiarezza reciproca deve essere al più alto livello e così la trasparenza: non è detto che sistemi di AI riescano a gestire una tale complessità e ambiguità.
Anche perché errori di giudizio sono a volte collegati a dati reali, quali il diverso modo di organizzare e gestire i propri armamenti nucleari. Ad esempio, per uno stratega russo, la logica della strategic stability, alla base degli accordi per il controllo degli armamenti strategici (SALT, START) non si applica alle armi nucleari tattiche, che a suo avviso contribuirebbero solo indirettamente alla deterrenza nucleare, come le armi convenzionali. Questa non è la logica della deterrenza allargata americana nella NATO, per la quale ogni arma nucleare è comunque parte dell'equilibrio strategico: più che di armi tattiche si parla in questo caso di sistemi sub-strategici. Esistono molte altre differenze analoghe che potrebbero portare a disastrosi errori di percezione.[6]

L'uso dell'AI potrebbe creare pericolose illusioni strategiche. Così ad esempio non mancano coloro che ritengono che la capacità dell'AI di processare e integrare una immensa quantità di dati potrebbe consentire di programmare un "primo colpo" nucleare devastante, capace di distruggere preventivamente anche le capacità nucleari di "secondo colpo" dell'avversario, facendo così saltare una volta per tutte il principio della Mutual Assured Destruction che è alla base dell'equilibrio della deterrenza nucleare.

In realtà anche solo le contromisure già esistenti, che potrebbero facilmente essere potenziate, sembrano in grado di impedire un tale rischio, ma molto dipende dalla

[6] Thomas C. Shelling, *La diplomazia della violenza*, Quaderni IAI, Il Mulino, Bologna 1968. Robert Jervis, *Perception and Misperception in International Politics*, revised edition, Princeton University Press, Princeton 1976

percezione della realtà da parte dei maggiori attori, ed è possibile che questa venga distorta da una eccessiva fiducia nel progresso tecnologico.[7]

Ma come evolveranno gli equilibri internazionali con il crescere di importanza dei sistemi di AI? E saranno sempre gli stati e i governi nazionali a definirne le caratteristiche o vedremo l'affermarsi di nuovi attori, quali ad esempio le imprese che detengono e sviluppano queste tecnologie o magari nuove aggregazioni socio-politiche (nuove "tribù") fidelizzate attraverso le loro interazioni con diversi sistemi di AI? O magari si formerà una sorta di sistema misto, un collage o un patchwork che raccolga insieme vecchi e nuovi centri di potere, in un equilibrio tutto da definire? Un articolo di Kissinger, assieme a due grandi esperti della materia (un ex presidente di Google e il fondatore di una impresa cyber ed ex consulente di Microsoft), pone questi ed altri interrogativi sulle possibili conseguenze politiche e strategiche dello sviluppo dell'AI. Le possibili evoluzioni potrebbero rivoluzionare l'intero sistema internazionale come oggi lo conosciamo, ad esempio rimettendo in discussione il cosiddetto ordine di Westphalia e l'affermarsi degli stati nazionali, per riproporre nuove forme "imperiali" di sfere di influenza. O potrebbe superare il concetto stesso di guerra. Ma la questione di fondo è quale sarà il ruolo dell'uomo, dell'umanità, in un tale futuro, che potrà essere pacifico o terribilmente distruttivo, e se riuscirà ancora a controllarlo ed indirizzarlo.[8]

Questi e molti altri problemi richiederebbero un approccio politico internazionale, una ripresa del dialogo sul controllo degli armamenti, centrato sul rapporto con l'AI. Come sostiene un recente studio della RAND è necessario come minimo stabilire un ecosistema che consenta agli uomini e all'AI di coesistere pur mantenendo la capacità umana di scegliere e di affermare i propri valori, a livello individuale e sociale. Inoltre è necessario costruire un'architettura globale di sicurezza che possa mantenere l'equilibrio tra gli attori che svilupperanno l'ecosistema. Bisogna insomma negoziare un accordo su alcuni principi di base validi per ogni sistema di AI.[9]

---

[7] Vedi anche San Winter-Levy and Nikita Lalwani, *The End of Mutual Assured Destruction? What AI Will Mean for Nuclear Deterrence*, Foreign Affairs, August 2025.

[8] Henry A. Kissinger, Eric Schmidt e Craig Mundie, *War and Peace in the Age of Artificial Intelligence*, Foreign Affairs, November 18, 2024

[9] Vedi Joel B. Predd, James Baker, Benjamin Boudreaux, Edward Geist e Matt Chessen, *Finding Common Ground in AGI Strategy Debates*, RAND Expert Insights, marzo 2026, Santa Monica (CA)._-

ISSN 2038-1662

Non è detto però che sia possibile trovare un accordo, o anche solo un linguaggio comune tra i vari (forse troppi) centri che in qualche modo prendono le decisioni relative all'AI. Un recente studio pubblicato da Chatham House constata che i protagonisti non sono sulla stessa lunghezza d'onda. USA e Cina puntano ognuno alla superiorità nazionale. Le medie potenze cercano a fatica di colmare un divario che invece si allarga. Le istituzioni nazionali e internazionali non hanno il potere di controllare gli sviluppi dell'AI né spesso la competenza tecnica. Gli investimenti privati superano di gran lunga quelli pubblici. Alleati o rivali, spesso non trovano un consenso neanche sulle basi di partenza. Un accordo è a questo stadio molto probabilmente impossibile.

Date queste premesse, lo studio suggerisce che l'opportunità per arrivare ad un accordo di governance efficace dell'AI potrebbe presentarsi col sopravvenire di qualche grande crisi sistemica, come spesso è avvenuto in passato per altri regimi (nucleari, finanziari, cyber, eccetera). Bisognerebbe quindi prepararsi a cogliere una tale opportunità preparando da subito un elenco delle azioni auspicabili e degli strumenti necessari, magari cominciando con lo stabilire una serie di canali di comunicazione tra coloro che si troverebbero a dover gestire l'eventuale crisi[10].

Sempre in questo spirito, un altro studio della RAND, valuta come ridurre i rischi che un così veloce sviluppo tecnologico può provocare e che uno Stato, agendo singolarmente, potrebbe non essere in grado di controllare. In particolare lo studio considera le conseguenze che scelte compiute da singoli responsabili di sistemi di AI (stati, imprese, gruppi o anche individui) potrebbero avere sul resto del mondo e si pone il problema di come continuare a promuovere l'innovazione, condividere i benefici dell'AI e al contempo proteggere gli interessi nazionali: come gestire i rischi connessi a questi sviluppi?

Il tema è come espandere la cooperazione internazionale tra gli stati e le altre entità transnazionali coinvolte. A tal fine lo studio analizza gli approcci di un gran numero

[10] Rowan Wilkinson, Alex Krasodomski, Isabella Wilkinson e Francisco Javier Varela Sandoval, *Breaking the Deadlock on AI Governance - How a Crisis Could Lead to Global Coordination*, The Royal Institute of International Affairs - Chatham House, London March 2026

di organizzazioni internazionali alla ricerca di funzioni chiave che potrebbero essere oggetto di una cooperazione strategica sull'AI. L'obiettivo è il perseguimento di tre finalità tra loro interrelate: 1) migliorare la comprensione delle capacità, delle implicazioni e dei rischi dell'AI, 2) promuovere lo sviluppo e l'utilizzo di sistemi AI affidabili e gestire la proliferazione di sistemi AI ad alta capacità, 3) mitigare e reagire alle conseguenze negative derivate da AI. Ciò richiederà probabilmente una reale collaborazione strategica internazionale quanto meno nel settore della ricerca AI, della determinazione di standard, del monitoraggio e della verifica. Altre funzioni possono essere utilmente prese in considerazione quali la formulazione di leggi e regole, la consultazione degli stakeholders, la condivisione delle informazioni, l'analisi delle prospettive di sviluppo futuro eccetera.[11]

La necessità di un approccio politico è maggiore da che i sistemi di AI sono divenuti "generativi" nel senso che sviluppano il proprio autonomo sistema di apprendimento. Non solo questo rende più importante l'inserimento nel sistema di alcuni elementi valoriali chiave (e la possibilità/necessità di rivederli ed integrarli nel tempo) ma è necessario prendere in considerazione la possibilità che, interagendo con altre AI, il sistema possa assumere profili imprevisti. Dopo tutto, diversi regimi politici stabiliscono diversi limiti ai loro sistemi specie per quel che riguarda il controllo sociale e politico. Allo stesso tempo, la lite scatenatasi negli USA tra il DoW e uno dei suoi più importanti fornitori di AI, proprio sui limiti di utilizzo del sistema rivela la totale mancanza di chiare linee guida: è possibile ed accettabile che a stabilire i limiti di utilizzo e i criteri morali e politici di riferimento di un sistema AI sia la ditta produttrice o anche un operatore ministeriale? Non è possibile un sistema più in linea con i principi di uno stato di diritto?

[11] Brodi Kotila, Katherine H. Tucker, Samantha Cherney, Austin Wyatt, *Strategic Cooperation on AI - Core Functions*, RAND Corporation, Santa Monica (CA), 2026

ISSN 2038-1662