Archivio Disarmo
IRIAD
Istituto di ricerche internazionali

USPID
Unione Scienziati per il Disarmo

Rapporto di ricerca settembre 2025

Lo stato dell'Intelligenza Artificiale in ambito militare e le prospettive di regolazione a livello nazionale, europeo e internazionale

Realizzato con il contributo del
Ministero degli Affari Esteri e della Cooperazione Internazionale

Lo stato dell'Intelligenza Artificiale in ambito militare e le prospettive di regolazione a livello nazionale, europeo e internazionale

Il presente Rapporto è stato redatto dall'Istituto di Ricerche Internazionali ARCHIVIO DISARMO in collaborazione con USPID - Unione degli scienziati per il disarmo. Esso è opera di un gruppo di ricerca diretto da Fabrizio Battistelli e formato da Guglielmo Tamburrini (cap. 1), Gian Piero Siroli (cap. 2), Diego Latella (cap. 3), Alice Saltini (cap. 4), Violetta Pagani (cap. 5), Alessia De Benedictis e Maurizio Simoncelli (cap. 6), Fabrizio Battistelli e Francesca Farruggia (cap. 7).

Si ringraziano Giorgia Pelosi e Matteo Taucci per la cura editoriale.

Il progetto è stato realizzato con il contributo del Ministero degli Affari Esteri e della Cooperazione Internazionale ai sensi dell'art. 23 bis del D.P.R. 18/1967.
"Le posizioni contenute nel presente report sono espressione esclusivamente degli autori e non rappresentano necessariamente le posizioni del Ministero degli Affari Esteri e della Cooperazione Internazionale".

Roma, settembre 2025

Lo stato dell'Intelligenza Artificiale in ambito militare e le prospettive di regolazione a livello nazionale, europeo e internazionale

Indice **Pag.**

Acronimi

A differenza delle sigle che indicano istituzioni (ONU, NATO, UE ecc.), le sigle che designano entità o funzioni specifiche sono riportate in corsivo (p. es. *DIU*: *Diritto Umanitario Internazionale*; *LAWS*: *Lethal Autonomous Weapon Systems*; ecc.).

AA = Armi Autonome
ADM = Armi di Distruzione di Massa
AUV = *Autonomous Underwater Vehicle* (Veicolo Autonomo Subacqueo)
AWS = *Autonomous Weapon Systems* (Sistemi d'Arma Autonomi)
CCA = *Collaborative Combat Aircraft* (Velivolo da Combattimento Collaborativo)
[C]CCW = [*Convention on*] *Certain Conventional Weapons* ([Convenzione su] Certe Armi Convenzionali)
CEC = Consiglio Ecumenico delle Chiese
CEND = *Creating an Environment for Nuclear Disarmament initiative* (Iniziativa per la Creazione di un Clima Favorevole al Disarmo Nucleare)
CICR = Comitato Internazionale della Croce Rossa
CIO = *Chief of Information Officer* (Responsabile delle Informazioni)
CNSS = *Committee on National Security Systems* (Comitato per i Sistemi di Sicurezza Nazionale, USA)
COTS = *Commercial-Off-The-Shelf*
CSSP = *Cybersecurity Service Providers* (Fornitori di Servizi di Sicurezza Informatica)
CUS = Controllo Umano Significativo
DC3 = *Department of Defense Cyber Crime Center* (Centro per la Criminalità Informatica del Dipartimento della Difesa, USA)
DCSA = *Defense Counterintelligence and Security Agency* (Agenzia per la Sicurezza e il Controspionaggio del Dipartimento della Difesa, USA)
DIANA = *Defence Innovation Accelerator for the North Atlantic* (Acceleratore dell'Innovazione della Difesa per l'Atlantico Settentrionale)
DIB = *Defense Industrial Base* (Base Industriale della Difesa, USA)
DIU = Diritto Internazionale Umanitario
DoD = *Department of Defense* (Dipartimento della Difesa, USA)
DoS = *Department of State* (Dipartimento di Stato, USA)
DSB = *Defense Science Board* (Consiglio Scientifico della Difesa, USA)
GAO = *Government Accountability Office* (Ufficio per la Rendicontazione del Governo, USA)
GEG = Gruppo di Esperti Governativi
GRIP = *Groupe de Recherche et d'Information sur la Paix et la Sécurité* (Gruppo di Ricerca e di Informazione sulla Pace e la Sicurezza)
HLEG = *High Level Expert Group on Artificial Intelligence* (Gruppo di Esperti di Alto Livello sull'Intelligenza Artificiale)
IA = Intelligenza Artificiale

ICRAC = *International Committee for Robot Arms Control* (Comitato Internazionale per il Controllo delle Armi Robot)
IDF = *Israeli Defense Forces* (Forze di Difesa Israeliane)
IOT = *Internet of Things*
ISC = *International Student Conference* (Conferenza Internazionale degli Studenti)
ISR = *Intelligence, Surveillance and Reconnaissance* (Intelligence, Sorveglianza e Ricognizione)
IT = *Information Technology* (Tecnologia dell'Informazione)
JIMS = *Joint Incident Management System* (Sistema Congiunto di Gestione degli Incidenti)
LAWS = *Lethal Autonomous Weapon Systems* (Sistemi d'Arma Letali Autonomi)
LLM = *Large Language Models* (Modelli Linguistici di Grandi Dimensioni)
MAD = *Mutual Assured Destruction* (Distruzione Reciproca Assicurata)
NC3 = *Nuclear Command Controll and Comunications systems* (Sistemi di Comando, Controllo e Comunicazione Nucleare)
NIST = *National Institute of Standards and Technology* (Istituto Nazionale per la Metrologia e la Tecnologia, USA)
NNSA = *National Nuclear Security Administration* (Amministrazione Nazionale per la Sicurezza Nazionale, USA)
NPT = *Non-Proliferation Treaty* (Trattato di Non Proliferazione)
NTCDL = *Network Tactical Common Data Link* (Rete Tattica Comune di Collegamento dei Dati)
NW-IT = *Nuclear Weapons Information Technology* (Tecnologia dell'Informazione sulle Armi Nucleari)
OMB = *Office of Management and Budget* (Ufficio di Gestione e Bilancio, USA)
OT = *Operational Technology* (Tecnologia Operazionale)
REAIM = *Responsible Artificial Intelligence in the Military Domain* (Intelligenza Artificiale nell'Ambito Militare)
RUSI = *Royal United Services Institute, UK*
SCADA = *Supervisory Control and Data Acquisition Systems* (Sistemi di Controllo, di Supervisione e di Acquisizione dei Dati)
SD = Sistemi Digitali
SIGACT = *Significant Activity Report* (Rapporto sulle Attività Significative)
SIGINT = *SIGnal INTelligence* (Spionaggio di Segnali Elettromagnetici)
SKR = *Stop Killer Robots*
SSDIA = Sistemi Avanzati di Supporto alle Decisioni
STRATCOM = *United States STRATegic COMmand* (Comando Strategico degli Stati Uniti)
TD = Tecnologie Digitali
TEV & V = *Testing, Evaluation, Verification & Validation* (Test, Valutazione, Verifica e Convalida)
UNGA = *United Nations General Assembly (*Assemblea Generale delle Nazioni Unite*)*

UNSG = *United Nations Secretary General* (Segretario Generale delle Nazioni Unite)
USPID = Unione degli Scienziati Per il Disarmo
WASP-HS = *Wallenberg AI, Autonomous Systems and Software Program - Humanity and Society*
WFUNA = *World Federation of United Nations Associations* (Federazione Mondiale delle Associazioni delle Nazioni Unite)
WILPF = *Women's International League for Peace and Freedom* (Lega Internazionale delle Donne per la Pace e la Libertà)
XAI = *eXplainable Artificial Intelligence* (Intelligenza Artificiale Interpretabile)

Sommario

Il presente Rapporto, frutto della collaborazione tra l'Istituto di Ricerche Internazionali Archivio Disarmo e l'USPID - Unione degli Scienziati per il Disarmo, fornisce una ricognizione sistematica e approfondita dello stato dell'Intelligenza Artificiale (IA) in ambito militare, affrontando le sfide etiche, i rischi tecnici e strategici, nonché le prospettive normative e regolative a livello nazionale, europeo e internazionale.

In un momento storico in cui la corsa alla militarizzazione delle tecnologie emergenti si sviluppa in assenza di quadri giuridici vincolanti, questo studio intende contribuire a una riflessione critica e documentata sulle implicazioni della IA per la pace, la sicurezza internazionale e la tutela della dignità umana.

Il Rapporto si apre con una riflessione sui fondamenti etici dell'impiego della IA in ambito bellico, soffermandosi sul concetto di *Controllo Umano Significativo* (*CUS*), inteso come prerequisito per il rispetto del *Diritto Internazionale Umanitario* (*DIU*). La distinzione tra *Sistemi d'Arma Autonomi* (*AWS*) dotati di capacità decisionali e operative, e *Sistemi di Supporto alle Decisioni* (*SSDIA*), che influenzano le scelte degli operatori umani, rappresenta il punto di partenza per un'analisi che pone in discussione la possibilità di mantenere una supervisione effettiva sull'uso della forza. In entrambi i casi, infatti, le condizioni operative attuali – caratterizzate da elevata velocità decisionale, complessità ambientale e pressione cognitiva – minano le basi stesse del controllo umano, favorendo forme di "autonomizzazione" che rischiano di rendere le decisioni di vita o di morte opache, non verificabili e giuridicamente non attribuibili. Il Rapporto evidenzia come, anche in presenza di supervisione formale, i sistemi algoritmici possano indurre fenomeni di delega passiva, trasformando il controllo umano in un atto simbolico privo di efficacia sostanziale.

A queste problematiche si affianca l'analisi delle vulnerabilità tecniche dei sistemi militari basati sulla IA. Le applicazioni già operative – che spaziano dalla guerra elettronica alle operazioni cibernetiche, dalla sorveglianza e ricognizione all'analisi predittiva e alla logistica – espongono le Forze Armate a minacce multiple: manipolazione dei dati, attacchi *adversarial*, modelli opachi di tipo *black box*, comportamenti emergenti non previsti, disallineamento tra obiettivi umani e finalità algoritmiche. Il rischio non riguarda solo l'errore tecnico, ma l'intero equilibrio strategico. L'adozione di sistemi intelligenti nei meccanismi di Comando e Controllo può comprimere i tempi decisionali al punto da rendere impossibile l'intervento umano, alterando le logiche di deterrenza e aumentando la probabilità di escalation involontarie.

Particolarmente allarmante risulta il quadro tracciato in merito alla sicurezza informatica dei sistemi digitali militari. Attraverso un'analisi del contesto statunitense, il Rapporto mostra come molte infrastrutture critiche, incluse quelle nucleari, siano caratterizzate da architetture obsolete, protocolli di sicurezza inadeguati e vulnerabilità persistenti. I dati relativi agli incidenti informatici verificatisi tra il 2015 e il 2021 confermano l'urgenza di un cambio di paradigma, che preveda la sicurezza come requisito progettuale e non come elemento correttivo posteriore. La crescente interconnessione tra

IA, cyberspazio e armamenti strategici espone infatti i sistemi militari a minacce ibride e asimmetriche, amplificando il rischio di crisi incontrollabili.

In questo contesto, l'integrazione della IA nei *Sistemi di Comando, Controllo e Comunicazione Nucleare* (*NC3*) rappresenta una delle sfide più delicate. Il Rapporto analizza i possibili impatti dell'automazione sui processi decisionali ad altissimo rischio, mettendo in luce come l'impiego di modelli di apprendimento automatico in ambito nucleare, sebbene ancora in fase sperimentale, sollevi interrogativi profondi sulla sicurezza globale. Le criticità identificate riguardano la scarsa trasparenza dei processi decisionali, il fenomeno delle cosiddette "allucinazioni algoritmiche", la vulnerabilità a interferenze esterne, e infine la possibilità che la macchina persegua obiettivi divergenti da quelli umani, anche in assenza di malizia programmata.

Alla luce di queste criticità, il Rapporto dedica un'ampia sezione all'esame delle iniziative normative e diplomatiche attualmente in discussione. Viene ricostruito il dibattito in sede ONU, con particolare riferimento ai lavori del *Gruppo di Esperti Governativi* (*GEG*) nell'ambito della *Convenzione su Certe Armi Convenzionali* (*CCW*).

Sebbene la questione degli *AWS* sia oggetto di discussione da oltre un decennio, i negoziati hanno dovuto affrontare posizioni divergenti tra gli Stati partecipanti.

Il Rapporto identifica tre principali orientamenti negoziali. Su un fronte vi sono gli Stati "liberalizzatori", come Israele, Russia e Stati Uniti, che si oppongono a qualunque forma di regolazione vincolante e sostengono che il diritto internazionale esistente sia già sufficiente a disciplinare l'uso dei nuovi sistemi. Essi si dichiarano favorevoli all'autoregolazione da parte dei singoli Stati e al mantenimento della libertà tecnologica, rifiutando ogni ipotesi di trattato internazionale.

Sul fronte opposto si collocano gli Stati "proibizionisti", in prevalenza appartenenti al Sud globale – tra cui Argentina, Messico, Costa Rica, Egitto, Pakistan – che chiedono un bando totale degli *AWS*. Essi, infatti, giudicano inaccettabili questi sistemi sul piano etico e giuridico, poiché incompatibili con i principi fondamentali del diritto umanitario, in particolare in relazione alla distinzione tra combattenti e civili, alla proporzionalità e alla responsabilità.

Una terza posizione prende spunto dalla proposta del Comitato Internazionale della Croce Rossa (CICR), basata su un approccio "a doppio binario". Essa infatti prevede, da un lato, il divieto dei sistemi d'arma intrinsecamente incompatibili con il *Diritto Internazionale Umanitario* (*DIU*) – come quelli progettati per colpire esseri umani o caratterizzati da comportamenti imprevedibili – e, dall'altro, la regolamentazione rigorosa dei sistemi attraverso limiti spaziali, temporali, funzionali e operativi, nonché l'obbligo di mantenere un *Controllo Umano Significativo* (*CUS*).

Aderiscono alla posizione "dualista" Stati come Francia, Germania, Italia, Paesi Bassi ecc., i quali sostengono l'opportunità di vietare i sistemi più pericolosi e, al contempo, di regolamentare gli altri con vincoli giuridici precisi, promuovendo l'adozione di criteri condivisi e meccanismi efficaci di trasparenza, supervisione e attribuzione della responsabilità.

L’assenza di un consenso ha finora impedito l’adozione di uno strumento condiviso. Il principio del consenso che regola i lavori della *CCW* rappresenta un vincolo procedurale particolarmente rigido, che consente anche a una singola delegazione di bloccare qualsiasi avanzamento. Di fronte allo stallo delle trattative formali, si è progressivamente aperto il dibattito sulla possibilità di percorrere vie negoziali alternative, come il ricorso all’Assemblea Generale delle Nazioni Unite o l’avvio di iniziative regionali, per promuovere un accordo internazionale giuridicamente vincolante che stabilisca limiti chiari e universalmente applicabili all’impiego dell’Intelligenza Artificiale in ambito bellico.

Accanto all’azione diplomatica, il Rapporto evidenzia il ruolo cruciale svolto dalla società civile nella costruzione di una consapevolezza etico-politica e nella pressione per l’adozione di regole vincolanti. La Campagna internazionale *Stop Killer Robots* (*SKR)*, che raccoglie oltre 270 organizzazioni in 70 Paesi, ha assunto un ruolo di riferimento, promuovendo attività di *advocacy*, ricerca, educazione e mobilitazione pubblica. Comunità accademiche, attori culturali, chiese e reti religiose hanno contribuito ad ampliare il fronte della critica, proponendo una visione alternativa della sicurezza fondata sulla centralità della persona e sulla limitazione dell’autonomia letale delle tecnologie. Le posizioni espresse da Papa Francesco e da altre autorevoli voci internazionali sottolineano la dimensione morale e umana della questione, confermando la convergenza tra istanze etiche e richieste giuridiche.

A supporto di questo quadro, infine, il Rapporto presenta i risultati del sondaggio su un campione rappresentativo di cittadini italiani appositamente effettuato da Archivio Disarmo nel 2025, che rileva l’atteggiamento dell’opinione pubblica del nostro Paese nei confronti della IA militare. I dati raccolti mostrano una forte opposizione all’impiego di armi autonome (74%), un’ampia preoccupazione per l’uso da parte di attori non statali (77%), e un’attenzione marcata per i rischi legati all’opacità decisionale e ai possibili danni ai civili. Solo una minoranza ritiene che il tema debba essere regolato dai parlamenti nazionali, mentre prevale una chiara preferenza per un approccio multilateralista, affidato a istituzioni sovranazionali come l’Unione Europea e l’ONU.

In conclusione, il Rapporto sottolinea la necessità urgente di una governance condivisa e vincolante della IA militare, basata su quattro pilastri fondamentali: il mantenimento del “Controllo Umano Significativo”, la trasparenza dei processi decisionali, l’attribuzione chiara delle responsabilità giuridiche e il coinvolgimento attivo della società civile. Solo attraverso una regolazione multilivello, fondata su principi etici condivisi, sarà possibile prevenire il rischio di conflitti automatizzati, deresponsabilizzati e probabilmente, in quanto delegati a “macchine”, anche più frequenti.

Abstract

This Report, the result of collaboration between the Istituto di Ricerche Internazionali Archivio Disarmo and USPID - Unione degli Scienziati per il Disarmo, provides a systematic and in-depth overview of the state of Artificial Intelligence (AI) in the military domain, addressing ethical challenges, technical and strategic risks, as well as regulatory and normative perspectives at the national, European, and international levels. At a historical moment in which the race to militarize emerging technologies is developing in the absence of binding legal frameworks, this study aims to contribute to a critical and well-documented reflection on the implications of AI for peace, international security, and the protection of human dignity.

The Report opens with a reflection on the ethical foundations of the use of AI in warfare, focusing on the concept of *Meaningful Human Control* (*MHC*), understood as a prerequisite for compliance with *International Humanitarian Law* (*IHL*). The distinction between *Autonomous Weapon Systems* (*AWS*) endowed with decision-making and operational capabilities, and *Decision Support Systems* (*DSS*), which influence the choices of human operators, represents the starting point for an analysis that questions the possibility of maintaining effective supervision over the use of force. In both cases, in fact, the current operational conditions – characterized by high decision-making speed, environmental complexity, and cognitive pressure – undermine the very foundations of Human Control, fostering forms of increasing autonomy that risk making life-or-death decisions opaque, unverifiable, and legally unaccountable. The Report highlights how, even in the presence of formal supervision, algorithmic systems can induce phenomena of passive delegation, transforming Human Control into a symbolic act devoid of substantive effectiveness.

Alongside these issues is the analysis of the technical vulnerabilities of military systems based on AI. The applications already in use – ranging from electronic warfare to cyber operations, from surveillance and reconnaissance to predictive analysis and logistics – expose the armed forces to multiple threats: data manipulation, adversarial attacks, opaque *black box* models, unforeseen emergent behaviours and misalignment between human objectives and algorithmic goals. The risk involves not only technical errors but the broader strategic balance. The adoption of intelligent systems in Command and Control mechanisms may compress decision-making timelines to the point of rendering human intervention impossible, thus altering deterrence logics and increasing the likelihood of unintended escalation.

Particularly alarming is the picture drawn regarding the cybersecurity of military digital systems. Through an analysis of the U.S. context, the Report shows how many critical infrastructures, including nuclear ones, are characterized by outdated architectures, inadequate security protocols, and persistent vulnerabilities. The data on cyber incidents that occurred between 2015 and 2021 confirm the urgency of a paradigm shift, in which security is seen as a design requirement and not as a subsequent corrective element. The growing interconnection between AI, cyberspace, and strategic weaponry

exposes military systems to hybrid and asymmetric threats, amplifying the risk of uncontrollable crises.

In this context, the integration of AI into *Nuclear Command, Control and Communication* (*NC3*) *Systems* represents one of the most delicate challenges. The Report analyzes the possible impacts of automation on extremely high-risk decision-making processes, highlighting how the use of machine learning models in the nuclear field, although still in the experimental phase, raises profound questions about global security. The identified criticalities concern the lack of transparency in decision-making processes, the phenomenon of so-called "algorithmic hallucinations", vulnerability to external interference, and finally the possibility that the machine may pursue objectives divergent from human ones, even in the absence of programmed malice.

In light of these concerns, the Report dedicates a broad section to the examination of current regulatory and diplomatic initiatives. The debate within the UN is reconstructed, with reference to the work of the *Group of Governmental Experts* (*GGE*) within the framework of the *Convention on Certain Conventional Weapons* (*CCW*).

Although the issue of *AWS* has been under discussion for more than a decade, negotiations have had to face divergent positions among participating States.

The Report identifies three main negotiating approaches. On one side are the "liberalizing" States, such as Israel, Russia, and the United States, which oppose any form of binding regulation and argue that existing international law is already sufficient to govern the use of new systems. They express support for self-regulation by individual States and the preservation of technological freedom, rejecting any proposal for an international treaty.

On the opposite front are the "prohibitionist" States, mostly from the Global South – including Argentina, Mexico, Costa Rica, Egypt, and Pakistan – which call for a total ban on *AWS*. They consider these systems ethically and legally unacceptable, as they are incompatible with the fundamental principles of humanitarian law, particularly with regard to the distinction between combatants and civilians, proportionality, and accountability.

A third intermediate position is inspired by the proposal of the International Committee of the Red Cross (ICRC), based on a "dual track" approach. It provides, on the one hand, for the prohibition of weapon systems that are inherently incompatible with International Humanitarian Law – such as those designed to target human beings or characterized by unpredictable behaviors – and, on the other hand, for the strict regulation of systems through spatial, temporal, functional, and operational limits, as well as the obligation to maintain *Meaningful Human Control*.

This "dualistic" position is supported by States such as France, Germany, Italy, and the Netherlands, which support the opportunity to prohibit the most dangerous systems while at the same time regulating the others through precise legal constraints, promoting the adoption of shared criteria and effective mechanisms for transparency, supervision, and attribution of responsibility.

The absence of consensus has so far prevented the adoption of a shared instrument. The consensus principle that governs the work of the *CCW* represents a particularly rigid procedural constraint, allowing even a single delegation to block any progress. In the face of the stalemate in formal negotiations, the debate has progressively opened to the possibility of pursuing alternative negotiation paths, such as turning to the United Nations General Assembly or initiating regional initiatives, to promote a legally binding international agreement that sets clear and universally applicable limits on the use of Artificial Intelligence in the military domain.

Alongside diplomatic action, the Report highlights the crucial role played by civil society in building ethical-political awareness and in pressing for the adoption of binding rules. The international *Stop Killer Robots Campaign* (*SKR*), which brings together over 270 organizations in 70 countries, has taken on a leading role, promoting *advocacy*, research, education, and public mobilization. Academic communities, cultural actors, churches, and religious networks have contributed to broadening the scope of criticism, proposing an alternative vision of security based on the centrality of the human person and on the limitation of the lethal autonomy of technologies. The positions expressed by Pope Francis and other authoritative international voices emphasize the moral and human dimension of the issue, confirming the convergence between ethical demands and legal requirements.

To support this framework, the Report also presents the results of a survey conducted by Archivio Disarmo in 2025 on a representative sample of Italian citizens, which records the attitude of public opinion in our country towards military AI. The collected data show strong opposition to the use of autonomous weapons (74%), widespread concern about their use by non-state actors (77%) and marked attention to the risks associated with decision-making opacity and possible harm to civilians. Only a minority believes that the issue should be regulated by national parliaments, while a clear preference emerges for a multilateralist approach, entrusted to supranational institutions such as the UN and the European Union.

In conclusion, the Report underlines the urgent need for shared and binding governance of military AI, based on four fundamental pillars: the maintenance of *MHC*, transparency in decision-making processes, clear attribution of legal responsibilities, and the active involvement of civil society. Only through multilevel regulation founded on shared ethical principles will it be possible to prevent the risk of automated, unaccountable conflicts which, because they are delegated to machines, may also become more frequent.

Introduzione

A differenza delle tecnologie consolidate (compresa quella nucleare) l'Intelligenza Artificiale (IA) è ancora a livello di progettazione e sperimentazione, tradizionalmente la fase più opaca e protetta di tutte. Se ciò è vero nel settore civile, lo è ancora di più, per ovvi motivi, nel settore militare. Viceversa, i rischi insiti nella IA per la guerra sono tanti e tali che, paradossalmente, è proprio nella fase della ricerca e sviluppo che sarebbe più opportuno e urgente ipotizzare e attuare misure di regolazione nella prospettiva dell'*Arms Control*. Se non verranno approntati interventi concordati, c'è motivo di prevedere che, in un futuro assai prossimo, più che di rischi si dovrà parlare di minacce.

Rinunciando, anche come esercizio di ottimismo, a utilizzare in questa sede i concetti di *rischi* e di *minacce* (Battistelli & Galantino, 2019), abbiamo aggregato le caratteristiche costitutive del fenomeno IA in quattro categorie, cui corrispondono altrettante parti del Rapporto. Ad esse abbiamo dato il nome di *sfide*: la sfida *etica*, la sfida *tecnologica*, la sfida *giuridico-diplomatica*, la sfida *sociale*.

Trattando di applicazioni belliche della IA, è necessario distinguere preliminarmente tra due ambiti e relativi "prodotti". Al primo ambito appartengono le *armi autonome* vere e proprie, quelle cioè che incorporano in sé sia la "mente", sia il "braccio": l'una attua il processamento dei dati e la conseguente selezione delle azioni da intraprendere; l'altro aziona il potenziale distruttivo (ordigni e munizioni) in grado di colpire l'obiettivo. Al secondo ambito appartengono i S*istemi di Supporto alle Decisioni* i quali, privi di capacità operativa, si avvalgono dei sistemi d'arma, suggerendo agli operatori o alle armi stesse la selezione e l'attacco degli obiettivi nemici (Amoroso, Mauri, Tamburrini, in corso di stampa).

Con una semplificazione logica e comunicativa, nella quale è preponderante il peso della cultura di massa, le armi autonome costituiscono la più nota delle applicazioni della IA, mentre i *Sistemi di Supporto alle Decisioni* possiedono oggi le potenzialità più importanti ed esercitano le funzioni più praticate (emblematico l'impiego degli algoritmi sui quali l'esercito israeliano basa le sue operazioni a Gaza). Invece è soltanto nell'immaginario collettivo (almeno per ora) che prevalgono l'allestimento e lo schieramento di falangi di macchine più o meno antropomorfe, destinate a combattersi fra loro (Farruggia, 2024).

Effettuata la necessaria distinzione circa l'autonomia rispettivamente dei sistemi d'arma e dei sistemi di supporto decisionale, la parte I del Rapporto, dedicata alla sfida etica, coincide con il capitolo 1 "Il problema del controllo umano sulle applicazioni belliche della IA".

Tutto il dibattito etico (cui si ricollega poi il dibattito giuridico) ruota intorno alla necessità che in entrambi i casi venga assicurato il *Controllo Umano Significativo* (*CUS*). Nato in relazione alle armi autonome, recentemente il problema si è riproposto anche con i sistemi di supporto decisionale, nonostante questi ultimi non siano dotati di capacità operativa diretta. La questione centrale è se, e come, sia possibile garantire una supervisione umana effettiva, in grado di prevenire violazioni del *Diritto Internazionale Umanitario* (*DIU)* e proteggere la dignità umana.

Nel caso delle armi autonome, il controllo umano (pre-condizione per il rispetto del *DIU*) rischia di diventare funzionalmente superfluo, se l'intero ciclo – dalla percezione del contesto alla selezione e attacco di un obiettivo – deve avvenire in totale autonomia. Difficilmente le armi autonome possono rispettare i principi del *DIU*: il principio di distinzione, che impone di colpire solo obiettivi militari legittimi; quello di proporzionalità, che vieta danni eccessivi ai civili rispetto al vantaggio militare ottenuto; e il principio di responsabilità, minato dall'opacità delle decisioni algoritmiche. Dal punto di vista etico, il rispetto della dignità umana impone che decisioni di vita o di morte non siano affidate esclusivamente a una macchina (CICR, 2021). Tuttavia, i recenti sviluppi tecnologici rendono l'intervento umano semplicemente impossibile nelle situazioni che nel gergo militare si definiscono "cinetiche", caratterizzate da alta velocità e/o da assenza di comunicazioni sicure. Questo scarto tra evoluzione tecnologica e capacità regolatoria è una delle criticità più gravi del momento attuale.

Analoghe problematiche si riscontrano nel caso dei sistemi di supporto. Pur non essendo autosufficienti, questi sistemi influenzano in modo significativo il processo decisionale umano. L'interazione operatore-macchina può produrre forme di delega passiva o di controllo nominale, in cui l'essere umano si limita ad approvare automaticamente i suggerimenti della macchina. Il caso documentato delle Forze di Difesa Israeliane (IDF), che hanno utilizzato sistemi di IA per generare liste di obiettivi nella Striscia di Gaza, mostra come la pressione operativa e l'aumento della produttività indotto dai sistemi di supporto possano compromettere l'esercizio attento e scrupoloso del controllo umano (Abraham, 2024).

Secondo gli studi di psicologia cognitiva, in condizioni di stress e urgenza, prevalgono processi decisionali intuitivi rispetto a quelli riflessivi. Equivoci come quelli indotti dall'euristica cosiddetta *WYSIATI* (*What You See Is All There Is*), oppure l'"ancoraggio al dato" provocato dall'algoritmo, aumentano il rischio di un'accettazione automatica delle decisioni della IA. Inoltre, la fiducia eccessiva nelle capacità della macchina, alimentata dal feticismo tecnologico, può inibire la responsabilità individuale per l'utilizzatore diretto e suscitare un'indebita pressione all'intensificazione delle operazioni assegnate al suo supervisore. Tra un approccio cognitivo analitico basato sul ragionamento e uno euristico basato sulla rapidità ed economicità del risultato, tende a prevalere il secondo. Per questo è necessario intervenire non solo sul design tecnico dei sistemi, ma anche sulle condizioni organizzative e psicologiche dell'interazione essere umano-macchina. I

sistemi di supporto devono essere trattati come dispositivi socio-tecnici, la cui affidabilità dipende da un contesto favorevole all'esercizio critico del controllo umano. Viceversa, la corsa alla militarizzazione della IA e alla compressione dei cicli decisionali rischia di trasformare il controllo umano in un rituale vuoto (Nadibaidze *et al.*, 2024).

In definitiva la tutela del controllo umano si conferma centrale per evitare derive disumanizzanti nell'impiego della IA in ambito militare. Nonostante le differenze che esistono tra armi da una parte e sistemi di supporto (entrambi "autonomi") dall'altra, le une e gli altri pongono problemi di rendicontabilità, di trasparenza e di rispetto del *DIU*. La proposta avanzata dalla Croce Rossa Internazionale e discussa nel Gruppo degli Esperti ONU[1] dei "due livelli" (la proibizione e le restrizioni) rimane un punto di riferimento utile, ma richiede aggiornamenti alla luce degli ultimi sviluppi tecnologici e delle dinamiche operative reali.

Delineati alcuni tra i principali problemi etici presentati dalle armi autonome, la parte II del Rapporto, comprendente i successivi tre capitoli, si concentra sui rischi presentati dalle applicazioni della IA in campo militare.

Il capitolo 2 "La IA nel dominio bellico: vulnerabilità e rischi" esamina le vulnerabilità tecniche, i rischi strategici e le implicazioni per la sicurezza internazionale insiti nelle applicazioni militari nella IA militare, una delle innovazioni strategiche più radicali e destabilizzanti dopo l'introduzione delle armi nucleari.

Nella crescente collaborazione tra l'essere umano e la macchina i problemi di imprevedibilità e di controllo si accrescono. Ad esempio, il supporto decisionale militare è sempre più assistito dalla IA con sistemi di simulazione, gestione della battaglia e analisi predittiva che aumentano la dipendenza da apparati opachi e indeboliscono la capacità critica degli operatori umani. In effetti, l'impiego della IA nella guerra elettronica e nell'*Intelligence, Sorveglianza e Ricognizione (ISR)* ha ampliato la capacità di raccogliere, analizzare e sintetizzare informazioni in tempo reale, ma comporta anche rischi non banali derivanti da falsi miti positivi, scelte automatizzate non trasparenti e decisioni letali affidate agli algoritmi (DoD, 2017).

Confermando la correlazione positiva che si stabilisce tra l'aumento della sofisticazione e l'incremento della vulnerabilità, i sistemi di IA militari risultano particolarmente vulnerabili agli attacchi avversari, alla corruzione dei dati (*data poisoning*), alla fragilità algoritmica e ai problemi di robustezza in ambienti complessi e non strutturati come quelli bellici. Sono frequenti i fenomeni di allucinazione algoritmica, disallineamento tra obiettivi e comportamenti appresi, sovra-parametrizzazione e scarsa capacità di adattamento a casi limite. La natura di *black-box* della IA, infine, rende

[1] V. oltre par. 5.3.

difficile la tracciabilità delle decisioni e la diagnosi degli errori, per non parlare della verifica, tanto cruciale quanto complessa da attuare, del rispetto del *DIU* e della conformità alle norme etiche e giuridiche in guerra.

È quindi urgente l'intervento della comunità internazionale per disegnare un quadro di governance multilaterale che affronti le minacce alla stabilità strategica e le lacune normative, specie nei termini della sicurezza umana. La cooperazione tecnica, il rilancio della prospettiva dell'*Arms Control* e la definizione di principi etici condivisi appaiono misure essenziali per prevenire un'escalation incontrollata. A tale scopo vengono formulate raccomandazioni che includono lo sviluppo di una "IA spiegabile", test avanzati di robustezza, sistemi di verifica delle decisioni e meccanismi internazionali per il monitoraggio delle applicazioni militari della IA. Prima che la rapidità di sviluppo di queste tecnologie di IA sopravanzino irrimediabilmente le capacità regolative e diplomatiche attuali.

Nel capitolo 3 "La sicurezza informatica nei sistemi militari" prosegue il tema delle vulnerabilità della IA in ambito militare, affrontato dal punto di vista della sicurezza delle tecnologie digitali e con particolare riferimento al contesto statunitense.

Un primo ambito di vulnerabilità riguarda i *Sistemi Digitali* (*SD*), spesso soggetti a malfunzionamenti o anomalie. Queste debolezze possono essere sfruttate da attacchi informatici capaci di compromettere l'integrità, la riservatezza e la disponibilità dei dati. Inoltre, vi sono rischi legati alla produzione e all'acquisizione dei sistemi, durante i quali attori ostili possono introdurre malware, aprendo *backdoor* sfruttabili in un secondo momento. Nonostante la possibilità di realizzare architetture resilienti mediante specifiche metodologie, per ragioni economiche il loro utilizzo rimane confinato ad ambiti ristretti. Invece, per i sistemi più diffusi, inclusi quelli basati sulla IA, la sicurezza viene spesso affrontata in maniera reattiva; quindi, *dopo* che le vulnerabilità sono state scoperte (Farrar, 2025).

Nell'emblematico caso delle infrastrutture militari statunitensi, numerosi test hanno mostrato la relativa facilità con la quale esse possono essere compromesse da attacchi anche poco sofisticati. Particolarmente delicato è il problema delle armi nucleari, la cui sicurezza informatica è affidata alla National Nuclear Security Administration (NNSA). L'integrazione crescente delle tecnologie digitali nei processi produttivi e nei sistemi di valutazione delle armi ha richiesto l'adozione di sei distinte pratiche di gestione del rischio, applicate ai tre ambienti: tradizionale, tecnologico-operativo, nucleare-militare. Da un'analisi del Government Accountability Office (GAO), il bilancio è allarmante: nessuno dei tre ambienti sopracitati presenta un'attuazione delle pratiche di sicurezza pienamente soddisfacente, dato che tra il 2015 e il 2021, il Dipartimento della Difesa (DoD) ha registrato oltre 12.000 incidenti informatici (GAO, 2022).

Insomma, la fragilità dei *SD* militari è tale da suscitare allarme, tanto più considerando il diretto impatto esercitato sulla stabilità strategica globale. La crescente automazione, la connessione in rete e l'uso massiccio della IA nei sistemi d'arma impongono un cambio di paradigma nella direzione di un impegno preventivo e sistemico, che integri la sicurezza fin dalla progettazione. Il caso delle armi nucleari è emblematico: senza una governance attiva, la vulnerabilità informatica diventa un rischio per la sicurezza globale.

Come logico sviluppo della sfida tecnologica, il capitolo 4 "La IA e i sistemi decisionali nucleari: rischi e implicazioni strategiche" approfondisce l'integrazione della IA nei sistemi decisionali nucleari, con particolare riferimento ai *Sistemi di Comando, Controllo e Comunicazione Nucleare* (*NC3*). Sebbene il tema sia ancora relativamente marginale nel dibattito pubblico e diplomatico, l'impiego della IA in tale contesto solleva interrogativi urgenti, in un quadro segnato da competizione strategica fra superpotenze, assenza di trasparenza e crisi degli accordi e delle regole condivise.

Dopo aver richiamato i rischi di escalation, errori sistemici e dipendenza eccessiva dagli output della IA, il capitolo propone un quadro analitico strutturato su tre fattori chiave: (a) le caratteristiche del modello di IA; (b) il grado di prossimità al processo decisionale; (c) il livello di supervisione e ridondanza umana. La combinazione di questi elementi determina l'alto o basso rischio delle singole integrazioni. L'adozione della IA negli *NC3* – già in fase di modernizzazione – può offrire vantaggi in termini di rapidità e resilienza decisionale, ma comporta anche una potenziale erosione dei meccanismi di controllo umano, in un ambiente dove non è tollerabile la fallibilità.

Dal supporto alla consapevolezza situazionale fino all'automazione dell'allerta precoce i modelli attuali – basati sul *deep learning* – si discostano radicalmente dagli algoritmi logico-regolativi della Guerra fredda, mostrando capacità predittive avanzate ma anche margini di errore e opacità di processo che li rendono inadeguati a funzioni decisionali ad alto rischio. Nonostante ciò, le simulazioni nella collaborazione tra grandi sviluppatori (come OpenAI o Anthropic) e i laboratori nucleari statunitensi confermano che le integrazioni non sono solo ipotesi speculative ma già oggetto di sperimentazione concreta (Chernavskikh, 2024).

Anche dal punto di vista tecnico, i modelli di IA odierni presentano problematicità strutturali. L'inaffidabilità si manifesta nel fenomeno delle cosiddette allucinazioni, cioè nella generazione di output equivoci o errati o non verificabili. L'opacità impedisce di comprendere i processi decisionali interni, mentre le vulnerabilità cibernetiche aprono la possibilità a interferenze esterne in sistemi altamente sensibili. Vi è poi il disallineamento, cioè il rischio che i modelli perseguano finalità divergenti da quelle umane, anche "in buona fede": come dimostra una simulazione in cui un modello ha invocato l'uso del nucleare "per amore della pace" (Rivera *et al.*, 2024).

Sul piano geopolitico, le dichiarazioni ufficiali ribadiscono la centralità del controllo umano nelle decisioni sull'impiego nucleare. Tuttavia, il trend in corso presso le principali potenze punta all'integrazione sempre più spinta della IA, anche grazie a partnership pubblico-private con grandi aziende tecnologiche. In assenza di un quadro regolatorio vincolante, la corsa al vantaggio tecnologico rischia di prevalere sul principio di precauzione. Dunque, il capitolo sottolinea l'urgenza di sviluppare meccanismi istituzionali, regole tecniche e strumenti di trasparenza per valutare e governare l'impatto della IA nei sistemi decisionali nucleari. Il futuro della stabilità strategica non dipenderà solo dalla deterrenza convenzionale, ma anche dalla capacità di prevenire integrazioni ad alto rischio che possano sfuggire al controllo umano o alimentare escalation incontrollate.

Se fino a qui il Rapporto ha esaminato le due grandi sfide poste ai decisori dalla IA bellica – quella etica consistente nel prevenire il sopravvento delle macchine sugli esseri umani e quella tecnologica (e operativa) consistente nel prevenirne le vulnerabilità e le disfunzioni – è arrivato il momento di dedicare spazio alle possibili risposte. Nella parte III del Rapporto la sfida è di natura giuridica e diplomatica (con un'evidente valenza anche politica) per quanto riguarda la possibile regolazione della IA militare (capitolo 5). Le sfide sono invece di natura prettamente sociale nella parte IV del Rapporto che esamina l'indispensabile apporto degli attori della società civile (capitolo 6) e dell'opinione pubblica (capitolo 7) alla consapevolezza e alla volontà politica dei decisori.

L'esigenza di porre dei limiti alla ricerca, allo sviluppo e in tempi brevi al possibile impiego di sistemi d'arma definitivamente autonomi dal controllo umano, viene affrontata dal capitolo 5 intitolato "La regolazione della IA militare tra autonomia degli Stati e proposte di accordi". L'accelerazione nello sviluppo di queste tecnologie è un processo che ha aperto, oltre a scenari bellici inediti, profondi interrogativi giuridici. Il rischio che delle macchine decidano in autonomia sulla vita o sulla morte di esseri umani rappresenta una frattura nei paradigmi del *DIU*.

Come noto, l'automazione nei sistemi d'arma procede lungo tre gradi: i sistemi "programmati" (*Human in the Loop*), dove è necessaria l'autorizzazione umana per colpire; quelli "supervisionati" (*Human on the Loop*), che agiscono autonomamente ma sotto supervisione; e quelli "autonomi" (*Human out the Loop*), privi di un controllo umano diretto. È proprio quest'ultima categoria a preoccupare di più, poiché rende estremamente difficile garantire il rispetto dei principi di distinzione, proporzionalità e necessità posti a fondamento del *DIU*.

Nel contesto giuridico multilaterale patrocinato dalle Nazioni Unite, la *Convenzione su Certe Armi Convenzionali* (*CCW*) e il relativo *Gruppo di Esperti Governativi* (*GEG*) costituiscono lo spazio istituzionale primario per il confronto tra Stati sul controllo della IA militare. Tuttavia, i colloqui promossi dal *GEG* si sono rivelati lenti e complessi, ostacolati da profonde divergenze. Tre orientamenti principali si confrontano: i

"liberalizzatori" (Stati Uniti, Russia, Israele), contrari a vincoli normativi e favorevoli a un'autoregolamentazione. I "proibizionisti" (Sudafrica, alcuni Paesi dell'America Latina e del Sud globale), che chiedono un divieto totale della IA militare priva di controllo umano. Infine, i "dualisti" (Francia, Germania, Italia, Paesi Bassi, ecc.), fautori di un approccio "a due livelli" che vieti i sistemi incompatibili con il *DIU* e regoli gli altri con precisi vincoli operativi.

Il nucleo intorno a cui si sviluppa il confronto è il concetto più volte ricordato di *CUS*, ovvero l'effettiva capacità dell'operatore umano di comprendere, supervisionare e intervenire sul funzionamento del sistema d'arma. Questo principio, pur ampiamente evocato, è ancora privo di una definizione condivisa. Nel 2021 il Comitato Internazionale della Croce Rossa (CICR) ha elaborato una proposta articolata in due richieste: vietare i sistemi imprevedibili o progettati per attaccare esseri umani; regolamentare gli altri, imponendo limiti spaziali, temporali e funzionali, garantendo sempre la possibilità di disattivazione o intervento umano. La proposta del CICR ha ottenuto ampio sostegno tra le organizzazioni della società civile, le agenzie delle Nazioni Unite e numerosi Stati. Essa riflette la crescente preoccupazione per la "disumanizzazione" della guerra e per l'opacità decisionale dei sistemi di IA, che rende difficoltosa l'attribuzione di responsabilità giuridiche in caso di violazioni del *DIU*.

Un ruolo molto importante è stato svolto dalle risoluzioni delle Nazioni Unite. Nella risoluzione 79/L.77 approvata nel dicembre 2024 dall'Assemblea Generale con 166 voti a favore, si invitano gli Stati a negoziare un trattato giuridicamente vincolante entro il 2026, sottolineando i rischi connessi alla IA militare: abbassamento della soglia dell'uso della forza, proliferazione incontrollata, escalation tecnologica, e perdita del controllo umano.

Anche nell'Unione Europea si va facendo largo l'esigenza di una regolamentazione multilaterale della IA militare che promuova la centralità del *CUS*. Un ostacolo oggettivo a un'effettiva regolazione (realizzata invece in ambito civile con il Regolamento UE 2024/1689) è rappresentato dal vincolo delle competenze in materia di difesa tuttora detenuta dai singoli Stati membri. Analogamente la NATO non si pronuncia sulle strategie nazionali in materia di IA e si limita ad auspicare l'interoperabilità tra alleati.

Il capitolo 5 si conclude evidenziando la posta in gioco: l'equilibrio tra innovazione tecnologica e rispetto dei diritti umani. Senza un'azione concertata, giuridicamente vincolante e capace di anticipare le implicazioni delle tecnologie emergenti, il rischio è che il dominio bellico venga affidato a opachi processi decisionali algoritmici. Viceversa, il principio del *CUS* appare come l'unico argine credibile alla disumanizzazione del conflitto, e il fondamento essenziale per una governance responsabile della IA militare.

Il capitolo 6, intitolato "L'impegno della società civile in Italia e all'estero: chiese, associazioni, movimenti (2020-2025)", ricostruisce il protagonismo crescente della

società civile globale nella contestazione dell'autonomizzazione del conflitto bellico attraverso la IA. In risposta alle cruciali implicazioni etiche, giuridiche e sociali delle armi autonome, si è consolidato un articolato fronte di mobilitazione composto da campagne internazionali, think tank, reti religiose, centri di ricerca e attivisti che, con strumenti e linguaggi diversi, hanno inciso sul piano politico, culturale e normativo.

Fulcro di questa mobilitazione è la Campagna internazionale *Stop Killer Robots* (*SKR*), che ha aggregato oltre 270 organizzazioni in 70 Paesi, ponendosi come punto di riferimento per un accordo internazionale che vieti l'uso di sistemi d'arma privi di un *Controllo Umano Significativo*. In questi anni, la Campagna *SKR* ha rafforzato le proprie attività di *advocacy*, lanciato iniziative parlamentari, promosso linee guida per i decisori politici e attivato micro-finanziamenti destinati a campagne nazionali e locali. Attorno a questa iniziativa si è consolidata una rete di attori istituzionali e sociali, come il *Future of Life Institute*, *Reaching Critical Will*, *UNA-UK* e – in Italia – l'Istituto di Ricerche Internazionali Archivio Disarmo e la Rete Italiana Pace e Disarmo, che hanno svolto un ruolo centrale nel sensibilizzare opinione pubblica e vertici istituzionali.

Particolarmente significativa è stata la convergenza con le comunità religiose, che hanno espresso una netta condanna dell'automazione letale. Papa Francesco, in particolare, ha definito le armi autonome "eticamente inaccettabili", riaffermando la centralità della dignità umana e l'insostituibilità del giudizio morale nei processi decisionali legati alla vita e alla morte. Le posizioni del Vaticano si sono intrecciate con quelle di altre confessioni cristiane e interreligiose, contribuendo a rafforzare la legittimazione etica della campagna globale.

Accanto all'azione politica, si è sviluppata un'intensa attività di produzione e divulgazione della conoscenza. Ricerche, report e analisi critiche – tra cui spiccano quelle di *Human Rights Watch* (2012), *Geneva Academy* e il CICR (2024), *RUSI* (2024), *Public Citizen* (2024), *info.nodes* (2025) e *Archivio Disarmo* - IRIAD (2020; 2023; 2024) – hanno alimentato il dibattito pubblico e istituzionale. Tali contributi si sono accompagnati a eventi pubblici, conferenze universitarie, simulazioni ONU per studenti e appuntamenti divulgativi come festival e workshop tematici. Insomma, in un contesto globale in cui l'innovazione tecnologica procede rapidamente, la voce della società civile rappresenta un fondamentale elemento di iniziativa e di stimolo, capace di orientare le scelte politiche e normative verso soluzioni più etiche e responsabili.

Il capitolo 7, "L'opinione pubblica italiana e l'autonomia delle armi", affronta il tema del rapporto tra opinione pubblica italiana e IA militare, sullo sfondo del dibattito sulla regolamentazione della IA militare a livello internazionale.

Dopo aver richiamato i sondaggi Ipsos in 28 Paesi, che mostrano tassi di opposizione alla IA militare mediamente superiori all'80%, il capitolo presenta i risultati dell'indagine demoscopica effettuata da Archivio Disarmo nel 2025. In Italia, solo il 15% degli intervistati si dichiara ben informato sul tema, ma la percentuale di opposizione all'uso

della IA militare raggiunge il 74%. Con una contrarietà che aumenta con l'età, in particolare tra gli over 65, probabilmente in relazione a un background culturale pacifista ereditato dal secondo dopoguerra.

La ricerca adotta un approccio sociologico, distinguendo tra individui "reattivi" – favorevoli a misure punitive e all'autodifesa armata – e "adattivi", inclini alla delega dell'uso della forza alle istituzioni, con i reattivi che tendono a essere lievemente più favorevoli all'impiego della IA militare. Complessivamente, la società italiana mostra avversione nei confronti di una IA militare completamente autonoma, anche nei casi in cui vengano ipotizzati vantaggi militari come la riduzione di perdite umane tra i soldati.

Più del 60% del campione si dichiara contrario all'uso della IA militare in ogni circostanza, segnalando la preoccupazione per la tutela della popolazione civile, che la maggioranza degli intervistati teme possa aumentare con l'impiego delle armi autonome. Seguono i timori relativi all'uso da parte di attori non statali (77%), al rischio di malfunzionamenti tecnici (75%) e all'opacità della catena di comando (72%). In particolare, quest'ultimo aspetto mette in discussione la possibilità di attribuire responsabilità giuridiche in caso di violazioni del diritto bellico, sollevando dubbi sull'efficacia del quadro normativo esistente.

Il tema della responsabilità è centrale: alla domanda su chi debba rispondere in caso di uccisioni ingiustificate da parte di IA militari, la maggioranza (61%) indica il governo dello Stato, seguito da militari e aziende produttrici. Per quanto riguarda l'operatività sul campo a richiesta di accountability umana fa appello alla centralità del *CUS* considerato cruciale non soltanto dagli scienziati ma anche dal pubblico.

Quanto al piano politico e diplomatico, gli intervistati si esprimono a favore di soluzioni multilaterali. Alla domanda su chi debba avere voce in capitolo nella regolamentazione della IA militare, solo il 15% indica i parlamenti nazionali. Il 64% ritiene invece che siano competenti organismi sovranazionali come l'Unione Europea, l'ONU e le agenzie per il disarmo. Questa preferenza per l'approccio multilaterale riflette la percezione del rischio di portata globale posto dalle applicazioni militari della IA e la consapevolezza che la soluzione può essere perseguita soltanto da norme condivise a livello internazionale.

L'ultimo capitolo evidenzia anche l'importanza dell'informazione sui temi strategici. L'ignoranza sul funzionamento e sugli effetti delle armi autonome non coincide con l'indifferenza. Al contrario, l'avversione verso l'automazione letale è tanto più forte quanto più l'opinione pubblica è esposta a narrazioni sulla centralità dell'essere umano nelle decisioni belliche. In questo senso, la comunicazione istituzionale e il ruolo della società civile risultano decisivi. Pur poco informata, l'opinione pubblica italiana

manifesta un orientamento etico chiaro: rifiuto dell'autonomia letale, richiesta di responsabilità umana, preferenza per una governance multilaterale.

Riferimenti bibliografici

Abraham, Y. (2024). *"Lavender": The AI machine directing Israel's bombing spree in Gaza*. +972/Local call. Disponibile a: https://www.972mag.com/lavender-ai-israeli-army-gaza/.

Amoroso, D., Mauri, D., Tamburrini, G. (in corso di stampa). Warfare at AI Speed and the Extended Meaningful Hu-man Control Problem. *In Proceedings of the XXII Amaldi Conference.* Accademia dei Lincei, Roma, 28-30 novembre 2024.

Battistelli, F., & Galantino, M. G. (2019). Dangers, risks and threats: An alternative conceptualization to the catch-all concept of risk. *Current sociology*, *67*(1), 64-78.

Chernavskikh, V. (2024). *Nuclear Weapons and Artificial Intelligence: Technological Promises and Practical Realities*. Stockholm International Peace Research Institute.

CICR - Comitato Internazionale della Croce Rossa. (2021). *ICRC position on autonomous weapons systems.* International Committee of the Red Cross.

DoD - US Department of Defense. (2017). *Project Maven to Deploy Computer Algorithms to War Zone by Year's End'.* Disponibile a: https://www.defense.gov/News/News-Stories/Article/Article/1254719/project-maven-to-deploy-computer-algorithms-to-war-zone-by-years-end/.

Farrar, O. (2025). *Understanding AI Vulnerabilities. As artificial intelligence capabilities evolve, so too will the tactics used to exploit them.* Harvard Magazine.

Farruggia, F. (2024). Oltre L'Immaginario. Quando l'AI va in guerra. *Im@go. A Journal of the Social Imaginary*, (23), 235-247.

GAO - US Government Accountability Office. (2022). *Enhanced Attention Needed to Ensure Cyber Incidents Are Appropriately Reported and Shared in DoD*. *Report to Congressional Committees*. Disponibile a: https://www.gao.gov/products/gao-23-105084.

Human Rights Watch & International Human Rights Clinic of Harvard Law School. (2012). *Losing Humanity: The Case Against Killer Robots*. Disponibile a: https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots.

IRIAD - Istituto di Ricerche Internazionali Archivio Disarmo. (2020). LAWS Lethal autonomous weapon systems. La questione delle armi letali autonome e le possibili azioni italiane ed europee per un accordo internazionale. Rapporto di ricerca realizzato con il sostegno del Ministero degli Affari Esteri e della Cooperazione Internazionale. *IRIAD Review. Studi sulla pace e sui conflitti*, *07-08*. Disponibile a: https://www.archiviodisarmo.it/view/K0Y2nX8-UWjKHNM9OQ83o96Kv0-oDTrYQW5IYIxM1dE/iriad-review-luglio-agosto.pdf.

IRIAD - Istituto di Ricerche Internazionali Archivio Disarmo. (2023). *Intelligenza artificiale: pace o guerra?*. Disponibile a: https://www.archiviodisarmo.it/intelligenza-artificiale-pace-o-guerra-comunicato-stampa.html.

IRIAD - Istituto di Ricerche Internazionali Archivio Disarmo. (2024). *"Intelligenza" delle macchine e follia della guerra: le armi letali autonome*. Disponibile a: https://www.archiviodisarmo.it/intelligenza-delle-macchine-e-follia-della-guerra-le-armi-letali-autonome.html.

Nadibaidze, A., Bode, I., & Zhang, Q. (2024). *AI in Military Decision Support Systems: A Review of Developments and Debates*. Odense DK: Center for War Studies.

Public Citizen. (2024). Deadly and Imminent: The Pentagon's Mad Dash for Silicon Valley's AI Weapons. Disponibile a: https://www.citizen.org/article/deadly-and-imminent-report/.

Rivera, J., Mukobi, G., Reuel, A., Lamparth, M., Smith, C., & Schneider, J. (2024). *Escalation Risks from Language Models in Military and Diplomatic Decision-Making*. arXiv.

RUSI - Royal United Services Institute. (2024). *The Proliferation Risk of Lethal Autonomous Weapons – Paper Launch*. Rusi.org. Disponibile a: https://my.rusi.org/events/the-proliferation-risk-of-lethal-autonomous-weapons-paper-launch.html.

Parte I – La sfida etica

Cap. 1 – Il problema del controllo umano sulle applicazioni belliche della IA

1.1 Premessa

Le tecnologie della IA si stanno ampiamente diffondendo nel settore della difesa. In un ambito così vasto, varie applicazioni della IA vengono utilizzate per scopi che non riguardano finalità specifiche delle Forze Armate. Per esempio, ciò accade nel caso delle applicazioni della IA utilizzate per migliorare i servizi logistici, le catene di approvvigionamento di materiali, la gestione del reclutamento o delle carriere del personale militare. Ma le tecnologie della IA vengono utilizzate anche per sviluppare sistemi che svolgono funzioni specifiche delle Forze Armate. Vi sono, per esempio, sistemi di IA progettati per pianificare azioni militari, per compiere direttamente azioni sul campo di battaglia o per offrire supporto a decisioni da prendere in vari contesti bellici. Tra questi sistemi, i più noti a un pubblico più vasto di non specialisti sono verosimilmente i cosiddetti *Sistemi d'Arma Autonomi* (che indicheremo più brevemente con l'espressione Armi Autonome oppure con la sigla *AA*). Ciò dipende dal fatto che le *AA* hanno suscitato un ampio dibattito politico, etico e giuridico iniziato da circa due decenni. Il problema della regolamentazione delle *AA* si è sviluppato nell'ambito della *Convenzione su Certe Armi Convenzionali* (*CCW*) presso la sede delle Nazioni Unite di Ginevra nel corso dell'ultimo decennio, ed è approdato all'Assemblea Generale nel 2024 (Blanchard *et al.*, 2025). Le tecnologie della IA sono strettamente collegate alle *AA*, in quanto fattore abilitante fondamentale per il loro sviluppo e uso. Infatti, le *AA* più avanzate si basano su tecnologie della IA per l'apprendimento automatico nei settori della percezione artificiale, nella pianificazione e nell'esecuzione di azioni.

Oltre alle *AA*, le tecnologie della IA sono fondamentali anche per sviluppare *Sistemi Avanzati di Supporto alle Decisioni* (indicati, per brevità, con la sigla *SSDIA*), che elaborano informazioni da fornire come input al personale militare che ha il compito di prendere decisioni nelle varie fasi di pianificazione o svolgimento delle azioni belliche (Nadibaidze *et al*., 2024). Gli *SSDIA* sono generalmente sviluppati adottando metodologie di apprendimento automatico.

A differenza delle *AA*, gli *SSDIA* non hanno autonomia operativa in relazione a compiti di selezione e attacco di obiettivi militari. A dispetto di questa differenza cruciale, vi è un problema di carattere etico e giuridico che accomuna le *AA* e gli *SSDIA*. Il problema riguarda la richiesta di mantenere un *Controllo Umano Significativo* (*CUS*)[1] sulle operazioni di ambedue le tipologie di sistemi. Questo è il problema che viene esaminato in questo capitolo. La questione si è posta inizialmente in relazione alle *AA*, ma più recentemente è stata sollevata anche a proposito degli *SSDIA*:

- Bisogna mantenere un *CUS* su tali applicazioni della IA?

[1] Per una specificazione del termine "significativo", v. oltre par. 1.3.

- Se è così, in base a quali motivazioni etiche e giuridiche bisogna farlo?
- E in quale misura è possibile farlo?

Nell'affrontare questi interrogativi, si esporranno innanzitutto le motivazioni di natura etica o giuridica a favore del mantenimento del *CUS*. In base a tali motivazioni, si procederà poi a discutere se e come il *CUS* possa essere effettivamente mantenuto ed esercitato.

Il presente capitolo è organizzato come segue. Nel paragrafo 1.2 si riassumeranno le principali motivazioni etiche e giuridiche per il mantenimento del *CUS* sulle *AA*, le principali proposte che sono state messe in campo per la regolamentazione dell'uso delle *AA* che si ispirano a tali motivazioni etiche e giuridiche, nonché le difficoltà concrete per implementare tali proposte che emergono considerando alcuni recenti sviluppi tecnologici nell'ambito delle *AA* e di altri sistemi d'arma dotati di autonomia crescente nelle funzioni di selezione e attacco di un obiettivo.

Nel paragrafo 1.3, si metteranno in luce vari problemi di mantenimento del *CUS* sugli *SSDIA*. L'insorgere di questi problemi nel caso degli *SSDIA* può sembrare a prima vista paradossale, poiché tali sistemi sono esplicitamente progettati come ausili ai decisori umani, per fornire ad essi informazioni e suggerimenti nei processi decisionali, ma senza godere dell'autonomia operativa della quale in linea di principio possono godere le *AA*. Nonostante ciò, come vedremo, le condizioni di uso degli *SSDIA*, e in particolare le forme concrete dell'interazione essere umano-macchina in questo ambito, possono trasformare il *CUS* in un controllo che diventa puramente nominale. Nel caso di un controllo nominale, gli operatori umani non hanno la possibilità di vagliare con l'attenzione e le cautele dovute le proposte decisionali avanzate da un *SSDIA*. Si metterà in evidenza che il problema del *CUS* e di un suo possibile depotenziamento verso forme di controllo puramente nominale si pone con forza anche nel caso degli *SSDIA*, per quanto i sistemi di questo tipo non godano dell'autonomia operativa delle *AA*. Verranno esaminate le condizioni d'uso che si trovano alla radice di questa estensione del problema del *CUS* dalle *AA* agli *SSDIA*.

In ultima analisi, l'erosione del *CUS* sugli *SSDIA* e il conseguente passaggio a forme di controllo puramente nominali può essere indotto dalla competizione tra potenze militari a sfruttare i vantaggi potenziali offerti dalla velocità delle elaborazioni dei sistemi di IA. Tale velocità di elaborazione consentirebbe di accelerare il ritmo delle decisioni e delle azioni belliche tanto da assicurarsi una superiorità strategica e tattica sul nemico. Tuttavia, per ottenere tali vantaggi, le capacità cognitive, percettive e di reazione necessarie agli operatori umani che devono esercitare il *CUS* sugli *SSDIA* possono essere messe a dura prova. Il paragrafo 1.3 si chiude prendendo in considerazione varie contromisure per mitigare queste e altre minacce all'esercizio del *CUS* sugli *SSDIA*.

Nel paragrafo 1.4 si tirano le fila, anche passando brevemente in rassegna altri ambiti applicativi della IA che pongono sfide e problemi per il mantenimento della pace e il rispetto dei principi del *Diritto Internazionale Umanitario* (*DIU*) nei conflitti.

1.2. Nuovi sviluppi tecnologici e sfide per il controllo umano delle Armi Autonome

Il dibattito etico e giuridico sulle *AA* ha preso le mosse dall'identificazione delle principali proprietà funzionali delle *AA*. Secondo il Dipartimento della Difesa statunitense (DoD) affinché un sistema d'arma sia da considerarsi autonomo, esso deve essere capace di selezionare e attaccare un obiettivo senza richiedere ulteriori interventi da parte di operatori umani dopo la sua attivazione (DoD, 2012, pp. 13-14). Un'idea simile è stata espressa dal Comitato Internazionale della Croce Rossa (CICR). Secondo il CICR un'arma è autonoma solo se è in grado di selezionare e attaccare degli obiettivi "in modo indipendente" (CICR, 2016, p. 1). Per poter svolgere tali funzioni, è necessario che il sistema d'arma percepisca in qualche misura l'ambiente circostante, selezioni dei potenziali obiettivi all'interno del teatro bellico, compia delle scelte operative su quanto può fare, pianifichi e infine esegua le manovre di attacco. È perciò evidente che la IA costituisce una fonte primaria di conoscenze e metodologie per lo sviluppo delle *AA*, in quanto tecnologia abilitante per la percezione artificiale, la pianificazione e l'esecuzione di azioni.

Anche una descrizione informale e sommaria delle proprietà salienti delle *AA* consente immediatamente di comprendere perché lo sviluppo delle *AA* e la loro eventuale utilizzazione pongano un problema qualitativamente nuovo nella storia dell'interazione tra esseri umani e macchine in ambito bellico. La diffusione delle *AA* rende il controllo umano sull'azione bellica tecnicamente superfluo, poiché non si richiede l'inserimento di nessun operatore umano nel processo di elaborazione e decisione che conduce dalla selezione di un possibile obiettivo, alla pianificazione dell'attacco contro di esso e infine alla messa in atto dell'attacco. Si pone dunque il problema se da una prospettiva etica e giuridica risulti nondimeno necessario garantire il *CUS* – e cioè una supervisione umana efficace – sulle *AA*. In caso di una risposta affermativa, si pone un'ulteriore questione: se sia possibile esercitare il *CUS* sulle *AA* senza che si debbano privare completamente queste ultime delle loro funzioni distintive di selezione e attacco autonomo di un obiettivo. Si tratta di questioni che sono state ampiamente e lungamente dibattute – le origini del dibattito risalgono infatti a circa 20 anni fa – in ambito prima militare e accademico, poi diplomatico e intergovernativo.

A tutti questi livelli del dibattito, si è sviluppato nel corso dell'ultimo quinquennio un ampio consenso in merito alla necessità di conservare il *CUS* sulle armi autonome. Una importante piattaforma che delinea i modi in cui il *CUS* sulle *AA* debba essere esercitato è stata avanzata dal CICR nel 2021. Si tratta di una proposta "a due livelli", che contempla il divieto di alcune tipologie delle *AA* e restrizioni d'uso per altre. Per preparare il terreno a un'analisi di tale proposta a due livelli, proviamo innanzitutto a sintetizzare brevemente le soggiacenti motivazioni etiche e giuridiche[2].

Per prima cosa, sono stati sollevati dubbi in merito alla possibilità che le *AA* rispettino i principi della guerra giusta che sono stati poi ripresi e codificati dal *DIU*. La teoria della

[2] Per approfondimenti si rimanda a Tamburrini, 2020; Farruggia, 2023; Mecacci *et al*., 2024.

guerra giusta ammette che il ricorso alle armi sia moralmente giustificato in determinate situazioni di difesa da aggressioni belliche. Tuttavia, essa prescrive che tutte le parti coinvolte ottemperino a certi vincoli morali nella conduzione delle operazioni belliche. In particolare, va rispettata l'immunità dei non combattenti (Walzer, 1990). Questa richiesta è stata incorporata nel principio di distinzione del *DIU*, codificato nei protocolli aggiuntivi alla IV Convenzione di Ginevra del 1949[3].

Il principio di distinzione impone di limitare strettamente gli attacchi ad obiettivi militari che – per loro natura, ubicazione, scopo o uso – forniscono un contributo effettivo all'azione bellica e la cui distruzione, conquista o neutralizzazione offre un chiaro vantaggio militare nelle circostanze date (Protocollo I del 1977, art. 52). Per applicare correttamente il principio di distinzione è necessario discriminare tra i combattenti attivi e i nemici fuori combattimento, tra combattenti e popolazione civile inerme. È altresì necessario riconoscere e salvaguardare il personale militare, sanitario e religioso, le unità sanitarie e i mezzi di trasporto civili o militari.

Le caratteristiche degli ambienti bellici e dei campi di battaglia hanno sollevato dubbi a proposito della capacità di una *AA* di rispettare il principio di distinzione altrettanto bene di un soldato opportunamente addestrato. Un campo di battaglia non è quel mondo ben ordinato, ripetitivo e privo di sorprese che caratterizza una catena di montaggio robotizzata, dal quale sono state eliminate le possibili fonti di perturbazione dell'azione robotica. Si tratta piuttosto di un ambiente molto più dinamico e molto meno strutturato, nel quale gli attori coinvolti sfidano intenzionalmente la capacità di previsione degli avversari, mettendo in atto iniziative e manovre a sorpresa, azioni di hackeraggio, *jamming* e altri tentativi di perturbare il comportamento delle *AA*. Questa caratteristica degli ambienti operativi delle *AA* introduce notevoli difficoltà per la progettazione di test empirici che siano veramente adeguati a valutare la capacità delle armi autonome di affrontare correttamente situazioni impreviste che possono insorgere sul campo di battaglia (Cummings, 2021; Tamburrini, 2023).

Altri dubbi sono stati sollevati intorno alla capacità delle *AA* di rispettare il principio di proporzionalità, recepito dalla teoria della guerra giusta e codificato nel *DIU*. Questo principio impone di non sferrare un attacco che abbia un costo atteso eccessivo – in termini di vittime civili o di danni a installazioni civili – rispetto al vantaggio militare concreto e diretto che ne potrebbe derivare (Protocollo I del 1977, art. 51).

Il principio di proporzionalità richiede di bilanciare i vantaggi militari con i costi che il loro raggiungimento potrebbe comportare per la popolazione civile. Ma un bilancio preventivo di questo tipo presuppone l'uso di capacità cognitive ed emotive, di competenze sociali ed esperienze che non sono alla portata delle attuali tecnologie della IA in generale, e delle *AA* in particolare[4].

[3] Il testo è consultabile alla pagina: https://www.icrc.org/eng/assets/files/other/icrc_002_0321.pdf.
[4] Per una discussione più approfondita del problema si veda Amoroso *et al.*, 2021, pp. 76-96.

In estrema sintesi, le principali limitazioni delle attuali tecnologie della IA che ricorrono nelle argomentazioni etiche e giuridiche che esigono l'esercizio del *CUS* sulle *AA* sono le seguenti:

- Errori che possono rivelarsi costosi dalla prospettiva del *DIU* e che invece sono considerati ammissibili anche da valutazioni stringenti di accuratezza, condotte in fase di test, di un sistema di IA;
- La difficoltà di prevedere puntualmente, anziché solo statisticamente, il comportamento dei sistemi di IA;
- Limitata capacità dei sistemi di IA di adattarsi a contesti di azione mutevoli e poco strutturati come sono i campi di battaglia;
- Disallineamento parziale e difficilmente rilevabile tra le funzioni effettivamente apprese da un sistema di IA e i valori codificati nel *DIU* e che presiedono alla condotta militare *in bello* (Benjo *et al*., 2025);
- Debolezze delle capacità percettive e cognitive dei sistemi di IA in generale, e delle *AA* in particolare, che sono messe in evidenza da studi nel settore dell'*adversarial machine learning*. Tali debolezze possono risultare "sorprendenti" dalla prospettiva della psicologia cognitiva, in quanto normalmente non affliggono le capacità percettive e cognitive degli esseri umani (Szegedy *et al*., 2014)[5].

Un altro argomento a sostegno del *CUS* sulle *AA* non si basa su una considerazione di limitazioni o vulnerabilità che affliggono attualmente tecnologie e sistemi di IA. Questo argomento si incentra sul rispetto della dignità umana ed è indipendente dalle limitazioni o da eventuali progressi tecnologici nello sviluppo delle *AA*. In un rapporto stilato nella sua qualità di Relatore speciale delle Nazioni Unite per le esecuzioni extragiudiziarie, sommarie o arbitrarie, Christof Heyns ha messo in evidenza che la vittima designata di un'arma capace di prendere decisioni di vita o di morte non ha la possibilità di fare appello all'umanità condivisa di un "qualcuno" che si trovi dall'altra parte. Il valore intrinseco di un essere umano (ciò che gli conferisce dignità in quanto agente morale consapevole) sarebbe perciò negato a priori, qualora il suo destino fosse affidato alla discrezionalità delle scelte operate da una macchina, invece di comportare il giudizio e le responsabilità morali di un essere umano (Heyns, 2013).

Basandosi sul principio del rispetto della dignità umana, nonché sui principi di distinzione e proporzionalità incorporati nel *DIU* e sul loro possibile mancato rispetto da parte delle *AA*, il CICR ha elaborato nel 2021 uno schema generale per arrivare a un trattato internazionale vincolante che introduca proibizioni e restrizioni sullo sviluppo e l'uso delle *AA* (CICR, 2021). Lo schema contempla due richieste: divieto e regolamentazione restrittiva. La proposta del CICR si inserisce perciò in una famiglia più ampia di proposte di regolamentazione delle *AA*, che sono dette "a due livelli" in quanto comprendono sia divieti sia restrizioni. Bisogna sottolineare che le proposte di regolamentazione a due livelli riscuotono attualmente un consenso molto ampio nel

[5] Per un approfondimento su queste e altre limitazioni e vulnerabilità dei sistemi di IA, si rimanda a Tamburrini, 2023; v. oltre cap. 3.

dibattito internazionale sulle *AA* che si è sviluppato nell'ambito della *CCW* e, più recentemente, nel contesto dell'Assemblea Generale delle Nazioni Unite. Un consenso che si è rivelato essere molto più ampio rispetto a proposte iniziali più restrittive che prevedevano la totale messa al bando di qualsiasi *AA*[6].

La richiesta di divieto avanzata dal CICR riguarda le *AA* che siano progettate per attaccare direttamente le persone oppure che diano luogo ad effetti indiscriminati in violazione del principio di distinzione o del principio di proporzionalità del *DIU*. Le *AA* che non sono vietate su queste basi dovranno essere regolamentate nel loro impiego, rispettando una serie di vincoli che articolano più specificamente la richiesta generale di esercitare il *CUS* su tutti i sistemi d'arma e, in particolare, sulle *AA*.

Ecco, in sintesi, le due richieste individuate dal CICR, che trovano oggi un ampio consenso all'interno della comunità internazionale degli Stati come punto di partenza per lo sviluppo di accordi internazionali sulla regolamentazione delle *AA*.

1. Bisogna proibire le *AA* progettate o utilizzate in maniera tale che i loro effetti non possono essere previsti, compresi e spiegati. Questa proibizione ha lo scopo di prevenire effetti indiscriminati che possono scaturire dall'uso delle *AA* e che risultano essere incompatibili con il *DIU*. Bisogna inoltre proibire le *AA* che sono state progettate o vengono utilizzate per attaccare direttamente le persone. Il motivo principale di questa ulteriore proibizione è la salvaguardia della dignità delle potenziali vittime di un'azione bellica. Queste ultime non dovrebbero mai essere soggette a decisioni prese esclusivamente da una macchina e che riguardano la loro incolumità fisica;
2. Bisogna regolamentare mediante opportuni vincoli la progettazione e l'utilizzazione delle *AA* che non ricadono nel caso 1. In particolare, bisogna limitare a oggetti di natura esclusivamente militare il tipo di obiettivi che le *AA* possono attaccare. A questo scopo, bisogna introdurre limitazioni sulla durata, sull'area geografica e sulla portata dell'azione dell'arma autonoma, al fine di rendere possibile la supervisione umana su ogni specifico attacco. Bisogna inoltre introdurre vincoli sulle forme di interazione essere umano-macchina, per garantire che vi sia una supervisione umana efficace non solo in fase di progettazione e test, ma anche in corso d'opera. Questi vincoli devono segnatamente comprendere la possibilità che degli operatori umani siano in grado di intervenire tempestivamente per correggere le azioni di una *AA* oppure per disattivarla completamente.

Delle considerazioni etiche e giuridiche a sostegno del punto 1 si è già detto in riferimento alla teoria della guerra giusta, al *DIU* e al rispetto della dignità umana. Il punto 2 si basa sull'ipotesi che le *AA* sufficientemente prevedibili e comprensibili nei loro effetti esistano o possano essere progettate, in modo tale da consentire un'adeguata forma di *CUS* sul loro impiego per permettere l'attribuzione chiara di responsabilità a persone

[6] Sull'approccio "a due livelli", detto anche "dualista", e sull'adesione ad esso di numerosi Paesi, soprattutto europei, v. oltre cap. 5.

inserite nella catena di Comando e Controllo, qualora si verifichino violazioni di norme morali o giuridiche sulla condotta delle azioni belliche.

Un *Sistema d'Arma Autonomo* che sembra conformarsi a quanto richiesto al punto 2 è il sistema mobile antimissile *Iron Dome*, utilizzato in Israele per monitorare e neutralizzare con il lancio di missili intercettori razzi e altri proiettili balistici diretti verso il territorio israeliano[7]. Gli operatori posizionano sul terreno *Iron Dome* e circoscrivono l'area che esso deve monitorare e proteggere. Dopo aver compiuto queste operazioni preliminari, *Iron Dome* viene abilitato a rispondere in piena autonomia, lasciando però agli operatori la facoltà di disabilitare il sistema in corso d'opera. È dotato di analoghe capacità di intercettazione anche il sistema tedesco *Nbs Mantis*, utilizzato dalle Forze Armate tedesche per proteggere i soldati e le installazioni militari da proiettili balistici in arrivo[8].

Bisogna rilevare a questo punto che gli attuali sviluppi tecnologici delle *AA* (e di loro precursori) non sembrano sempre accordarsi con le richieste avanzate dal CICR. Ricordiamo, a questo proposito, che nel 2023 l'aviazione militare statunitense ha condotto con successo delle prove sperimentali su un aereo da caccia, la cui navigazione aerea è stata interamente controllata da un sistema della IA. Il caccia è stato testato in uno scenario di duello aereo a distanza ravvicinata (*dogfight*) con un altro caccia, eseguendo manovre di attacco e di evasione. I piloti che erano presenti nella cabina di pilotaggio del caccia interamente controllato da un sistema di IA avevano la facoltà di escludere il sistema, assumendo essi stessi il controllo della navigazione aerea. Ma nei test eseguiti non c'è mai stato bisogno che i piloti umani subentrassero[9]. Dai test condotti nello scenario di confronti aerei a distanza ravvicinata risulta evidente la possibilità tecnologica di trasformare questo aereo dotato di capacità autonome di navigazione e manovra in una vera e propria *AA*, "semplicemente" aggiungendovi le capacità di selezionare e attaccare un obiettivo.

Consideriamo ora questo recente sviluppo tecnologico dalla prospettiva assunta nel punto 2 della proposta avanzata dal CICR. È opportuno chiedersi a questo riguardo se il comportamento di un tale velivolo da combattimento sia sufficientemente prevedibile e ammetta le forme di supervisione e intervento umano contemplate dal CICR allo scopo di correggerne l'azione oppure di disattivarlo. Non è scontato che questa possibilità sia invariabilmente garantita in confronti aerei ravvicinati, nei quali potrebbe anche essere coinvolto un numero elevato di velivoli impegnati in interazioni cooperative o competitive tra loro. Inoltre, se i segnali di controllo sono impartiti a grande distanza dal velivolo – come accade attualmente nel caso dei droni militari, che possono essere pilotati da una stazione a terra situata a migliaia di chilometri di distanza – la latenza prolungata dei segnali di controllo non consente l'intervento tempestivo di operatori umani a scopi di correzione dell'azione o disattivazione. Sarebbe perciò necessario che i segnali di

[7] Per approfondimenti si rimanda alla pagina: https://www.army-technology.com/projects/irondomeairdefencemi/.
[8] Per approfondimenti si rimanda alla pagina: https://www.army-technology.com/projects/mantis/.
[9] Per approfondimenti si rimanda alla pagina: https://www.defensenews.com/air/2024/04/19/us-air-force-stages-dogfights-with-ai-flown-fighter-jet/.

controllo per correggere le manovre rapide del velivolo da combattimento siano impartiti da una postazione collocata a distanza più ravvicinata. E tuttavia, nelle missioni che comportano il sorvolo di territorio nemico o di zone militarmente contestate, questa possibilità di controllo sembra restringersi a postazioni collocate su altri velivoli sufficientemente vicini, con tutte le limitazioni del caso, che comprendono sia la perturbazione delle comunicazioni nello spettro elettromagnetico da parte del nemico sia la neutralizzazione delle postazioni ravvicinate di controllo.

In questi vari scenari, la possibilità di esercitare il *CUS* sul velivolo da combattimento si riduce o si annulla completamente, in contrasto con quanto prevede la condizione 2 della proposta di regolamentazione avanzata dal CICR. In relazione a tali scenari bisogna anche chiedersi quale possa essere il ruolo del principio di necessità militare incorporato nel *DIU*. Come verrà interpretato tale principio se il controllo ravvicinato verrà meno? Si rinuncerà a concludere la missione oppure si lascerà piena autonomia all'aereo da caccia senza pilota a bordo? Non è per nulla evidente che si possa conciliare l'opzione di lasciare piena autonomia all'aereo da caccia senza pilota con la proposta di regolamentazione del CICR, soprattutto nel caso in cui le missioni nelle quali si deciderà di impiegare il caccia autonomo non saranno solo di *dogfight*, ma eventualmente comprendano – introducendo opportune modifiche nel sistema di controllo del velivolo – anche missioni di bombardamento e di attacco di obiettivi selezionati al suolo.

In conclusione, la sperimentazione di un caccia per il *dogfight* e le eventuali evoluzioni di tale prototipo (che sono evidentemente alla portata delle attuali tecnologie della IA) entrano in tensione con il rispetto di vincoli eticamente e giuridicamente motivati contenuti nella piattaforma "a due livelli" del CICR per la regolamentazione delle *AA*. La possibilità tecnologica di un *dogfight* "autonomo" non era stata ancora avvalorata sperimentalmente nel 2021, all'epoca in cui fu resa pubblica la proposta del CICR. Ma è opportuno notare a tale proposito che la proposta è stata reiterata pressoché *verbatim* anche nel 2024 all'interno del contributo del CICR al Rapporto del Segretario Generale dell'ONU sulle "armi letali autonome" (UNSG, 2024a).

In conclusione, la perdurante assenza di una regolamentazione internazionale, a distanza di più di quattro anni dalla proposta del CICR, non può che aggravare il problema posto da uno sviluppo incalzante di tecnologie della IA per le *AA*, e cioè di una erosione progressiva del controllo degli esseri umani sulle *AA*, con le conseguenti minacce per il rispetto del *DIU*. Lo scarto tra esigenza normativa e realtà tecnologica in rapida evoluzione mette impietosamente in luce i ritardi delle attuali iniziative della società civile, degli organismi intergovernativi e della diplomazia sulla regolamentazione delle *AA*.

1.3. Sistemi di Supporto alle Decisioni Belliche e Controllo Umano Significativo

I problemi evidenziati nella sezione precedente in relazione all'esercizio del *CUS* sulle *AA* si propagano in larga misura, come ora vedremo, anche all'esercizio del *CUS* sugli *SSDIA*. Vi sono pertanto radici comuni per il mantenimento del *CUS* in questi due ambiti

distinti, per quanto gli *SSDIA* non abbiano l'autonomia operativa delle *AA* nello svolgimento di compiti di attacco di obiettivi militari.

Gli *SSDIA* in ambito militare fanno parte di una classe molto più ampia di sistemi di IA che svolgono funzioni di supporto al processo decisionale umano, anche in situazioni eticamente sensibili, nelle quali le decisioni prese e le azioni conseguenti hanno un impatto significativo sulla vita e sugli interessi delle persone. Questi ambiti spaziano dalla diagnosi medica (Kumar *et al*., 2023) alla pianificazione di interventi chirurgici (Ficuciello *et al*., 2019) fino agli interventi dell'amministrazione pubblica, quando si deve decidere, per esempio, se concedere dei servizi sociali o indagare il comportamento di alcuni contribuenti alla ricerca di frodi fiscali (Tan *et al*., 2023). I suggerimenti generati dagli *SSDIA* vengono attuati solo in modo condizionale, poiché sono valutati da operatori umani competenti e solo in base a tale valutazione vengono accettati, rivisti o perfino ignorati. In questo processo di valutazione e di eventuale sbarramento consiste l'esercizio del *CUS* sugli *SSDIA*. Anche il termine "significativo" che compare nell'espressione *Controllo Umano Significativo* (*CUS*) risale originariamente ai primi dibattiti etici e giuridici sulle *AA*. È stata opportunamente introdotta in quel contesto problematico per distinguere tra un controllo umano nominale da un lato – che comporta controlli sommari e superficiali – e un controllo umano attento, coscienzioso e approfondito dall'altro lato (Article 36, 2013). Ma questa distinzione si estende anche ai modi di esercitare la funzione di filtro che gli esseri umani dovrebbero svolgere nella loro interazione con i *SSDIA*. Evidentemente, per assolvere coscienziosamente le responsabilità che riguardano l'implementazione dei suggerimenti di un *SSDIA*, si presuppone che gli operatori umani esercitino il *CUS*.

In ambito civile, la rilevanza normativa del *CUS* emerge chiaramente dai vari requisiti chiave elencati nelle linee guida etiche dell'Unione Europea per una IA affidabile (HLEG, 2018), che comprendono l'autonomia e la supervisione umana sui sistemi di IA che svolgono ruoli di supporto alle decisioni. Allo stesso modo, una serie di requisiti legati al *CUS* è elencata nell'AI Act dell'Unione Europea, soprattutto in relazione a domini applicativi – che vanno dalla fornitura di servizi pubblici e privati essenziali all'immigrazione e al controllo delle frontiere – che sono classificati come ad alto rischio (Gazzetta Ufficiale dell'Unione Europea, 2024).

La rilevanza normativa del *CUS* è altrettanto evidente nel contesto dei settori militari che non rientrano nell'ambito di applicazione della normativa UE. Si pensi, ad esempio, ai suggerimenti che gli *SSDIA* forniscono in relazione alla distribuzione delle unità militari sul campo di battaglia o alla selezione di obiettivi militari legittimi. L'attuazione dei suggerimenti forniti da tali sistemi può avere conseguenze significative per l'incolumità e la vita dei militari e dei civili coinvolti. In particolare, il giudizio umano può svolgere un ruolo fondamentale per rivedere o bloccare suggerimenti di un *SSDIA*, i quali potrebbero portare a violazioni del *DIU*. Rimangono valide in questo ambito le considerazioni sugli attuali limiti delle tecnologie della IA menzionate nella sezione precedente a proposito del problema del *CUS* sulle *AA*:

- Errori che possono rivelarsi costosi dalla prospettiva del *DIU* e che sono ammessi anche da valutazioni stringenti di accuratezza, condotte in fase di test, di un sistema di IA;
- La difficoltà di prevedere puntualmente, anziché solo statisticamente, il comportamento dei sistemi di IA;
- Limitata capacità dei sistemi di IA di adattarsi a contesti di azione mutevoli e poco strutturati come possono spesso essere i campi di battaglia;
- Disallineamento parziale e difficilmente rilevabile tra le funzioni effettivamente apprese da un sistema di IA e i valori codificati nel *DIU* e che presiedono alla condotta militare *in bello* (Benjio *et al.,* 2025);
- Debolezze delle capacità percettive e cognitive dei sistemi di IA in generale, e delle *AA* in particolare, che sono messe in evidenza da studi nel settore dell'*adversarial machine learning*. Tali debolezze possono risultare "sorprendenti" dalla prospettiva della psicologia cognitiva, in quanto normalmente non affliggono le capacità percettive e cognitive degli esseri umani (Szegedy *et al*., 2014).

Questi vari elementi sostengono anche le argomentazioni volte a esigere l'esercizio del *CUS* sugli *SSDIA* militari. Ma l'esercizio del *CUS* su questi sistemi deve tenere conto di altre limitazioni, che riguardano le capacità decisionali degli esseri umani e che possono emergere con particolare forza nelle interazioni tra questi e le macchine "intelligenti" della IA. A questo riguardo, cerchiamo ora di evidenziare la necessità di evitare situazioni di interazione essere umano-macchina in contesti bellici, all'interno dei quali il controllo umano si riduce a un superficiale esame dei suggerimenti della macchina, che potrebbe pertanto sfociare in una accettazione acritica dei suggerimenti ricevuti, senza che l'operatore metta in atto tutte le dovute cautele. A tale fine facciamo ricorso a uno scenario di utilizzazione degli *SSDIA* descritto in un'inchiesta giornalistica relativa alle IDF (*Israel Defence Forces*) e alle sue modalità di pianificazione delle operazioni militari a Gaza, messe in atto all'indomani del massacro di civili innocenti perpetrato da Hamas in territorio israeliano il 7 ottobre 2023.

Secondo un'indagine giornalistica condotta da +972/Local Call, le IDF hanno utilizzato questi sistemi di supporto decisionale abilitati dalla IA (*Habsora*, *Lavender*, *Where's Daddy?*) per generare liste di potenziali bersagli da bombardare nella Striscia di Gaza[10]. Membri della divisione *targeting* delle IDF – dei quali è stato protetto l'anonimato – hanno testimoniato che l'automazione di questo compito ha aumentato drasticamente il tasso di generazione di potenziali obiettivi da parte della medesima divisione, portandone il numero "da 50 obiettivi all'anno a 100 obiettivi al giorno". Una squadra di specialisti del *targeting* aveva il compito di esaminare ogni bersaglio suggerito, approvandolo o respingendolo sulla base di un controllo di legittimità, e dunque soprattutto in riferimento alle norme del *DIU*. In merito alla modalità di svolgimento di questo compito di filtraggio, un membro della stessa divisione ha dichiarato: "prepariamo gli obiettivi

[10] Per approfondimenti si rimanda alle pagine: https://www.972mag.com/lavender-ai-israeli-army-gaza/; https://www.theguardian.com/world/2023/dec/01/the-gospel-how-israel-uses-ai-to-select-bombing-targets.

automaticamente e lavoriamo secondo una check-list. Siamo davvero come in una fabbrica. Lavoriamo velocemente e non c'è tempo per approfondire l'obiettivo. La percezione è quella di essere giudicati in base al numero di obiettivi che riusciamo a generare".

Da queste dichiarazioni si evince una forte preoccupazione per l'impatto dell'incremento "produttivo" impresso dagli *SSDIA* e per le sue ripercussioni sul giudizio umano. Emerge inoltre la percezione di una pressione psicologica ad allineare le decisioni di filtraggio dell'operatore umano al ritmo produttivo della macchina (si lavora "come in una fabbrica"), ciò risulta dalla convergenza di due fattori: (1) la riduzione delle risorse temporali necessarie a vagliare approfonditamente i suggerimenti forniti dalla macchina in base a valutazioni di legittimità degli obiettivi ("lavoriamo velocemente e non c'è tempo per approfondire l'obiettivo"), e (2) i nuovi criteri che sarebbero stati adottati da superiori nella gerarchia militare, dando priorità all'aumento della produttività nel valutare le prestazioni nell'attività di filtraggio umano ("la percezione è quella di essere giudicati in base a quanti obiettivi riusciamo a generare"). Questi due aspetti spiccano per contrasto con le condizioni ideali di tempo a sufficienza e assenza di pressioni indebite che sarebbero necessarie a garantire un attento ed equilibrato esercizio del *CUS*. Inoltre, l'inchiesta giornalistica in questione afferma che i test di accuratezza eseguiti su tali sistemi hanno evidenziato un tasso di errore fino al 10% nell'identificazione corretta di obiettivi da colpire.

Si può invocare la ricerca di un vantaggio militare accelerando i tempi del ciclo osservazione-decisione-azione come una fonte plausibile della pressione psicologica subìta dai membri della divisione *targeting* delle IDF. Il principale collo di bottiglia che impedisce di cogliere vantaggi di questo tipo è rappresentato dal *CUS*. Si generano code più lunghe di potenziali obiettivi, data la velocità dei suggerimenti forniti dagli *SSDIA* in questione, che non vengono smaltite con la stessa velocità se si esercita un controllo scrupoloso di legittimità. Per mitigare il blocco, i superiori gerarchici incentivano un esame più rapido dei suggerimenti forniti dagli *SSDIA* e ripongono una fiducia nell'accuratezza di tali suggerimenti, che risulta essere eccessiva alla luce dei risultati poco soddisfacenti dei test di accuratezza. In effetti, da una prospettiva etica e giuridica, la richiesta di avvicinare la produttività degli operatori umani alla produttività della macchina appare particolarmente grave in funzione di test che indicano un tasso di errore fino al 10% nell'identificazione corretta di obiettivi da colpire.

Lo scenario descritto in questa inchiesta giornalistica offre importanti spunti di riflessione in merito alla difficoltà di esercitare e mantenere il *CUS* sugli *SSDIA* impiegati in fasi di pianificazione o esecuzione di azioni belliche. In assenza di varie condizioni al contorno, una configurazione di controllo *Human in the Loop* non è evidentemente sufficiente a garantire che gli operatori umani siano in grado di filtrare le proposte della IA in modo attento, coscienzioso ed efficace. Anche in assenza di negligenza da parte degli addetti al *CUS*, perturbazioni varie possono impedire loro di esercitare il *CUS* dall'interno dell'anello di controllo e di esprimere giudizi ponderati sui suggerimenti delle macchine.

La pressione psicologica esterna e le scarse risorse temporali sono fonti note e sperimentalmente comprovate di *bias*, pregiudizi e altre perturbazioni significative del comportamento degli esseri umani. Senza alcuna pretesa di esaustività, elenchiamo qui di seguito alcune possibili fonti di perturbazione psicologica del giudizio che possono incidere in siffatte condizioni d'uso di un *SSDIA*:

A. La magnificazione esagerata dei vantaggi della IA (il cosiddetto *hype* tecnologico) che si riscontra diffusamente nelle attuali narrative sulla IA, e che può indurre aspettative troppo elevate sulle capacità di tali sistemi;
B. Disincentivi a contestare i suggerimenti degli *SSDIA* che le organizzazioni militari possono introdurre;
C. L'insorgenza di *bias* euristici che può essere facilitata dalla concessione di finestre temporali eccessivamente ristrette per l'esercizio del *CUS*.

Sui punti A e B ci siamo già soffermati nei limiti di spazio a disposizione. Passiamo a considerare ora il punto C. Secondo i modelli duali del processo decisionale umano (Strack & Deutsch, 2015) ai processi di ragionamento ponderato si affiancano processi euristici di soluzione dei problemi che consentono spesso, ma non invariabilmente, di prendere decisioni soddisfacenti in tempi più rapidi. Si accettano più facilmente gli esiti di processi euristici decisionali senza passare per il vaglio di processi di ragionamento ponderato allorché scarseggino le risorse cognitive e temporali a disposizione. A questo proposito, Daniel Kahneman ha introdotto una ben nota distinzione semplificativa tra due "sistemi" cognitivi, che sussumono rispettivamente processi decisionali euristici oppure analitici (Kahneman, 2012). Il sistema 1 raccoglie i processi decisionali euristici più veloci. Il sistema 2 raccoglie i processi decisionali riflessivi e analitici, più lenti e maggiormente dispendiosi in termini di risorse cognitive e temporali. I due sistemi operano contemporaneamente, ma non sempre in modo cooperativo. Le opzioni più facilmente selezionabili dal sistema 1 possono essere accettate per default, soprattutto quando il tempo a disposizione per decidere che cosa fare è limitato, e gli sforzi concomitanti del sistema 2 richiedono troppo tempo per essere portati a compimento. Le opzioni presenti negli elenchi di possibili obiettivi da bombardare generati dagli *SSDIA* utilizzati dalle IDF sono facilmente disponibili, mentre potrebbe essere necessario un notevole sforzo cognitivo per rifiutare tali opzioni su basi razionali e proporre un'altra linea d'azione.

L'euristica *WYSIATI* (*What You See Is All There Is*), identificata e verificata sperimentalmente da Daniel Kahneman e Amos Tversky, può contribuire a spiegare questo effetto psicologico. L'euristica *WYSIATI* riflette, secondo Kahneman, una "notevole asimmetria tra il modo in cui la nostra mente tratta le informazioni attualmente disponibili e quelle che non abbiamo". Le persone tendono a concentrarsi su ciò che è chiaramente visibile, trascurando le informazioni che non sono note o che non sono attualmente recuperate dalla memoria: "le informazioni che non vengono recuperate (anche inconsciamente) dalla memoria – osserva Kahneman – potrebbero anche non esistere". Si salta alle conclusioni sulla base di questa asimmetria e si va a costruire "la migliore storia possibile che incorpora le idee attualmente attivate", avallando per default

la convinzione (spesso ma non sempre corretta) che solo queste informazioni siano rilevanti (Kahneman, 2012, p. 85). In condizioni decisionali caratterizzate da scarsità di risorse cognitive e temporali a disposizione, l'euristica *WYSIATI* può indurre le persone a concentrarsi su ciò che è chiaramente visibile e a trattare come inesistenti le informazioni che non sono note o non sono attualmente recuperate dalla memoria.

L'attivazione di *WYSIATI* in condizioni di pressione temporale non indica invariabilmente opzioni vantaggiose. La pressione del tempo può indurre gli operatori umani ad approvare i suggerimenti di *targeting* della IA resi prontamente disponibili dagli *SSDIA*. E ciò può accadere anche nel caso di sistemi che, come quelli delle IDF, hanno dimostrato sperimentalmente un margine di errore del 10%. Il filtraggio umano dei suggerimenti di *targeting* affetti da un tasso di errore così alto è fondamentale per evitare conseguenze gravi, in termini di violazioni del *DIU*, che potrebbero essere evitate mediante l'esercizio del *CUS* – conseguenze che possono comprendere la morte di individui fuori combattimento, di civili non combattenti e di altre persone legalmente protette[11].

Analoghe considerazioni si applicano ai *bias* che insorgono da altri procedimenti euristici. Oltre a quelli indotti dall'euristica dell'autorità, cioè dalla tendenza a conformarsi al comportamento di superiori gerarchici o di figure che giocano il ruolo di modelli comportamentali, ricordiamo anche il *bias* da "ancoraggio", e cioè la tendenza a prendere decisioni basandosi eccessivamente su un'informazione iniziale, detta àncora, anche quando essa non è accurata (Kahneman, 2012, pp. 137-138). Nel caso in questione l'àncora sulla quale il controllore umano si potrebbe agganciare sarebbe rappresentata dai suggerimenti errati degli *SSDIA* utilizzati dalle IDF.

In conclusione, gli *SSDIA* devono essere considerati come dei dispositivi socio-tecnici. Un ampio spettro di fattori psicologici e sociali contribuisce a determinare le modalità effettive di utilizzo di tali sistemi. Il mantenimento di adeguate condizioni psicologiche e sociali di contorno è fondamentale per ridurre il rischio di passare dal *CUS* a un controllo puramente nominale degli operatori umani. Vi sono dunque buone ragioni per affermare che i problemi che emergono in relazione al *CUS* sulle *AA* si propagano in larga misura all'esercizio del *CUS* sugli *SSDIA*. Vi sono radici comuni per il mantenimento del *CUS* in questi due ambiti distinti, nonostante il fatto che gli *SSDIA* non abbiano l'autonomia operativa delle *AA* nello svolgimento di compiti di attacco di obiettivi militari.

1.4. Osservazioni conclusive

Le considerazioni svolte in questo capitolo sul problema di conservare nelle mani degli esseri umani le responsabilità delle azioni belliche sono rilevanti per uno sviluppo ulteriore del dibattito accademico, diplomatico e politico sulla militarizzazione della IA e sulla necessità di conservare il *CUS* sulla guerra allo scopo di evitare situazioni

[11] Per un'analisi di profili di responsabilità giuridicamente rilevanti che emergono in tali circostanze d'uso a carico di superiori gerarchici, si rimanda a Amoroso, 2024; Mauri, 2024.

accidentali di innesco o di escalation dei conflitti armati causati dal ricorso alle *AA* o agli *SSDIA*. In relazione alle *AA*, è stato evidenziato come il rapido sviluppo tecnologico ponga nuove sfide alle ipotesi di soluzione "a due livelli" per la regolamentazione delle *AA*.

In relazione agli *SSDIA*, sono stati messi in evidenza soprattutto problemi legati alla corsa alla militarizzazione della IA con la finalità di accorciare i cicli di osservazione-decisione-azione per acquisire vantaggi strategici e tattici sugli avversari sul campo di battaglia. Il principale rischio connesso a questa competizione tra potenze militari riguarda la trasformazione del *CUS* sugli *SSDIA* in un controllo nominale, nel quale gli operatori umani sono ridotti al ruolo di semplici passacarte, incapaci di vagliare e filtrare opportunamente i suggerimenti offerti dagli *SSDIA*, e posti in questa condizione da condizioni d'uso che non rispettano i limiti delle capacità cognitive e percettive degli esseri umani.

Pertanto, il contributo offerto in questo capitolo, oltre ad aggiornare il dibattito sulla regolamentazione delle *AA* alla luce di recenti sviluppi tecnologici, consente di approfondire anche tematiche relative ad applicazioni militari della IA in campi diversi dalle *AA*. L'opportunità di approfondire la riflessione nella direzione di applicazioni militari della IA in campi diversi dalle *AA* è sottolineata nella recente risoluzione 79/239 *Artificial intelligence in the military domain and its implications for international peace and security*, adottata dall'Assemblea Generale delle Nazioni Unite il 24 dicembre 2024 (UNSG, 2024b)[12].

Esulano dai limiti di quanto discusso in questo capitolo altre sfide che emergono dai rapidi e pervasivi sviluppi della IA e che riguardano l'obiettivo fondante delle Nazioni Unite di mantenere la pace e adoperarsi per la risoluzione pacifica dei conflitti. Sono compresi in questo orizzonte più ampio l'uso delle applicazioni degli *Large Language Models (LLM)* per scopi di intelligence militare, sicurezza e sorveglianza[13] nonché gli sviluppi tecnologici che pongono nuove sfide al rispetto dei trattati internazionali sul divieto di produzione, stoccaggio e uso di *Armi di Distruzione di Massa* (*ADM*).

Ricordiamo a questo proposito gli usi duali di applicazioni civili della IA per lo sviluppo di *ADM*. Un esempio notevole di *dual-use* è stato fornito da un gruppo di ricerca operante nel settore farmaceutico (Urbina *et al*., 2022). Il gruppo di ricerca ha mostrato che un sistema di IA, normalmente utilizzato per scoprire nuovi farmaci, può essere trasformato in un sistema per scoprire agenti chimici tossici – compiendo così un primo passo verso la produzione di *ADM*. Il sistema era stato originariamente addestrato per individuare nuovi composti chimici utili per la sperimentazione di nuovi farmaci. Nel corso dell'addestramento automatico che ha portato allo sviluppo di questo sistema, la tossicità per il corpo umano di un composto chimico veniva penalizzata, mentre veniva premiata la sua capacità di combattere vari agenti patogeni. Invertendo questa impostazione alla somministrazione di premi e penalità, il sistema è stato nuovamente

[12] Per un approfondimento sulla risoluzione 79/239 si rimanda a UNIDIR, 2025; v. oltre par. 5.4.

[13] Per approfondire si rimanda alla pagina: https://www.972mag.com/israeli-intelligence-chatgpt-8200-surveillance-ai/.

addestrato e premiato per l'individuazione di composti chimici altamente tossici per il corpo umano. Molte delle molecole identificate a seguito di tale nuovo processo di addestramento si sono rivelate più tossiche di agenti chimici tossici già noti (Urbina *et al*., 2022).

Il punto cruciale di questa storia è che un sistema di IA che promuove la salute umana e il diritto alla vita è stato agevolmente trasformato in un sistema per la costruzione di *ADM*. Emerge dunque evidente la necessità di un monitoraggio costante e di analisi delle molteplici e difficilmente prevedibili possibilità di applicazione della IA nell'ambito dei conflitti armati e dei problemi che tali sviluppi pongono rispetto al mantenimento della pace internazionale e al rispetto del *DIU* nei conflitti bellici (Grand-Clément, 2023). In questo capitolo è stato affrontato solo uno di tali problemi, legato al controllo umano delle *AA* e degli *SSDIA* sui campi di battaglia. Uno solo dei molteplici problemi posti dall'attuale corsa alla militarizzazione della IA; ma certamente non l'ultimo per ordine di importanza, in considerazione della competizione fra potenze militari per arrivare a combattere le guerre con l'efficienza e la velocità consentite dalla IA.

Riferimenti bibliografici

Amoroso, D. (ed.). (2021). *Autonomous weapons systems and international law. A study on human-machine interactions in ethically and legally sensitive domains*. Napoli: Baden-Baden - Edizioni scientifiche italiane Nomos.

Amoroso, D. (2024). Sistemi di supporto alle decisioni basati sull'IA e crimini di guerra: alcune riflessioni alla luce di una recente inchiesta giornalistica. *Diritti umani e diritto internazionale, 2*, 347–368.

Article 36. (2013). *Structuring debate on autonomous weapons systems. Memorandum for delegates to the Convention on Certain Conventional Weapons (CCW)*. Disponibile a: https://article36.org/wp-content/uploads/2013/11/Autonomous-weapons-memo-for-CCW.pdf.

Bengio, Y., Cohen, M., Fornasiere, D., Ghosn, J., Greiner, P., MacDermott, M., Mindermann, S., Oberman, A., Richardson, J., Richardson, O., Rondeau, M., St-Charles, P., & Williams-King, D. (2025). *Superintelligent agents pose catastrophic risks: Can scientist AI offer a safer path?*. arXiv.org.

Blanchard, A., Boulanin, V., Bruun, L., & Goussac, N. (2025). *Dilemmas in the policy debate on autonomous weapon systems*. Stockholm International Peace Research Institute.

CICR - Comitato Internazionale della Croce Rossa. (2016). *Views of the International Committee of the Red Cross on autonomous weapons systems*. Ginevra: International Committee of the Red Cross.

CICR - Comitato Internazionale della Croce Rossa. (2021). *ICRC position on autonomous weapons systems*. International Committee of the Red Cross.

Cummings, M. L. (2021). Rethinking the maturity of artificial intelligence in safety critical settings. *Artificial Intelligence Magazine, 42*, 6–15.

DoD - US Department of Defense. (2012). *Autonomy in Weapons Systems. US Department of Defense Directive 3000.09*. Disponibile a: https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf.

Farruggia, F. (a cura di). (2023). *Dai droni alle armi autonome*. Milano: FrancoAngeli.

Ficuciello, F., Tamburrini, G., Arezzo, A., Villani, L., & Siciliano, B. (2019). Autonomy in surgical robots and its meaningful human control. *Paladyn Journal of Behavioral Robotics, 10*, 30–43.

Gazzetta Ufficiale dell'Unione Europea. (2024). *AI Act. Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence*. Disponibile a: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.

Grand-Clément, S. (2023). *Artificial Intelligence Beyond Weapons: Application and Impact of AI in the Military Domain*. Ginevra: UNIDIR.

Heyns, C. (2013). *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions*. United Nations, Document A/HRC/23/47.

HLEG - High level expert group on artificial intelligence. (2018). *Ethics Guidelines for Trustworthy Artificial Intelligence*. Disponibile a: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

Kahneman, D. (2012). *Thinking, Fast and Slow*. London: Penguin Books.

Kumar, Y., Koul, A., Singla, R., & Ijaz, M. F. (2023). Artificial intelligence in disease diagnosis: A systematic literature review, synthesizing framework and future research agenda. *Journal of Ambient Intelligence and Humanized Computing, 14*(7), 8459–8486.

Mauri, D. (2024). Numeri, persone, umanità: sistemi di supporto alle decisioni umane in campo militare da parte dell'IDF e diritto internazionale umanitario. *Diritti umani e diritto internazionale, 2*, 329–346.

Mecacci, G., Amoroso, D., Cavalcanti-Siebert, L., Abbink, D., van den Hoven, J., & Santoni De Sio, F. (2024). *Research Handbook on Meaningful Human Control of Intelligent Systems*. Cheltenham UK: Edward Elgar.

Nadibaidze, A., Bode, I., & Zhang, Q. (2024). *AI in Military Decision Support Systems: A Review of Developments and Debates*. Odense DK: Center for War Studies.

Strack, F., & Deutsch, R. (2015). The duality of everyday life: Dual-process and dual system models in social psychology. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (eds.), *APA Handbook of Personality and Social Psychology, vol. 1. Attitudes and Social Cognition*. pp. 891–927. New York: American Psychological Association.

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2014). *Intriguing Properties of Neural Networks*. arXiv.org.

Tamburrini, G. (2020). *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*. Roma: Carocci.

Tamburrini, G. (2023). Il dibattito etico sulle armi autonome. In F. Farruggia (a cura di), *Dai droni alle armi autonome. Lasciare l'apocalisse alle macchine?*. pp. 98–112. Milano: FrancoAngeli.

Tan, E., Petit, J. M., & Simonofski, A. (2023). Artificial intelligence and algorithmic decisions in fraud detection: An interpretive structural model. *Data & Policy, 5*.

UNIDIR - United Nations Institute for Disarmament Research. (2025). *AI in the military domain: A Briefing Note for States*. Disponibile a: https://unidir.org/publication/ai-military-domain-briefing-note-states/.

UNSG - United Nations Secretary-General. (2024a). *Lethal autonomous weapons systems*: *Report of the Secretary-General*. Disponibile a: https://digitallibrary.un.org/record/4059475?v=pdf.

UNSG - United Nations Secretary-General. (2024b). *Artificial intelligence in the military domain and its implications for international peace and security: budget implications of draft resolution A/C. 1/79/L*. Disponibile a: https://digitallibrary.un.org/record/4065062?v=pdf.

Urbina, F., Lentzos, C., Invernizzi, C., & Ekins, S. (2022). Dual use of artificial-intelligence-powered drug discovery. *Nature Machine Intelligence, 4*, 189–191.

Walzer, M. (1990). *Guerre giuste e ingiuste. Un discorso morale con esemplificazioni storiche.* Trad. it. Napoli: Liguori editore.

Parte II – La sfida tecnologica

Cap. 2 – La IA nel dominio bellico: vulnerabilità e rischi

2.1. Premessa

Questo capitolo analizza l'evoluzione delle applicazioni della IA nella guerra moderna, concentrandosi sulle vulnerabilità emergenti e sui rischi strategici. L'integrazione crescente della IA nelle operazioni militari, nei processi decisionali e nelle armi autonome, altera il carattere e i paradigmi tradizionali della guerra, presentando sfide di sicurezza senza precedenti. Questo cambiamento rappresenta uno degli sviluppi tecnologici militari più significativi dall'avvento delle armi nucleari, con profonde implicazioni per la sicurezza internazionale e la stabilità strategica.

I principali risultati evidenziano che, sebbene la IA offra ovvi vantaggi militari, introduce anche vulnerabilità critiche come malfunzionamenti, manipolazioni ostili, lacune etiche e rischi di escalation. Il rapporto identifica cinque categorie di rischio che richiedono attenzione immediata: vulnerabilità tecniche, rischi nei sistemi di Comando e Controllo, impatto sulla stabilità strategica, minacce asimmetriche e problematiche di proliferazione.

Le caratteristiche uniche della IA includono la ipervelocità del progresso tecnologico, la natura *dual-use* della ricerca, l'accessibilità delle tecnologie della IA a vari attori, il funzionamento autonomo e l'imprevedibilità dei sistemi di apprendimento complessi.

Con queste caratteristiche ed il rapido sviluppo si scavalcano i quadri normativi esistenti, creando vuoti di sicurezza. La comunità internazionale, le Nazioni Unite e il diritto internazionale sottolineano la necessità urgente di cooperazione per sviluppare salvaguardie tecniche e nuovi approcci al controllo degli armamenti, instaurare fiducia tra gli attori in gioco e stabilire quadri etici condivisi e strumenti giuridicamente vincolanti. La finestra per un intervento diplomatico efficace verso una governance internazionale delle applicazioni militari della IA si sta restringendo, la consapevolezza e la capacità di intervento su questi temi è fondamentale.

"*Success in creating AI would be the biggest event in human history. Unfortunately, it might also be the last, unless we learn how to avoid the risks*".
Stephen Hawking, fisico e matematico (Hawking *et al*., 2014)

2.2. Stato attuale delle applicazioni militari della IA

Lo stato attuale delle applicazioni militari della IA è un campo in rapidissima evoluzione e dalle molteplici sfaccettature. La IA viene integrata in vari aspetti delle operazioni militari, dai sistemi autonomi agli strumenti decisionali. Ecco alcune aree di interesse:

2.2.1. Sistemi d'Arma Autonomi

Le applicazioni militari della IA si sono evolute significativamente al di là dei semplici sistemi automatizzati, includendo piattaforme sempre più autonome in grado di selezionare i bersagli ed ingaggiarli con vari gradi di supervisione umana. Questi sistemi vanno da piattaforme difensive con necessaria autorizzazione umana ad armi completamente autonome progettate per operare in modo indipendente anche in assenza di canali di comunicazione. Ad esempio, sistemi come il drone *Kargu-2* hanno dimostrato capacità come il puntamento autonomo e il riconoscimento facciale (ADF, 2025). Gli sviluppi principali includono:

- Munizioni circuitanti (*loitering munitions*) con capacità di riconoscimento autonomo dei bersagli;
- Sistemi di allerta per la protezione di confini e installazioni;
- Veicoli autonomi senza pilota nei domini aereo, terrestre e marittimo, dotati della IA per navigazione, riconoscimento dei bersagli e attacchi di precisione. Esempi di veicoli aerei sono l'MQ-9 *Reaper* e il *Velivolo da Combattimento Collaborativo* (*CCA*) XQ-58 *Valkyrie*, che rappresentano un cammino evolutivo verso la collaborazione essere umano-macchina (*Manned-Unmanned Teaming*) (Airforce Technology, 2024);
- Tecnologie di sciame che consentono operazioni autonome coordinate e l'interazione tra sistemi con o senza equipaggio, con condivisione e integrazione delle informazioni;
- Operazioni subacquee con *Veicoli Autonomi Subacquei* (*AUV*) per compiti quali il rilevamento di mine e la mappatura e sorveglianza di ambienti marittimi;
- Sistemi contro-IA progettati per sconfiggere le piattaforme autonome avversarie.

Mentre la piena autonomia rimane controversa, relativamente al concetto di *CUS*, le barriere tecniche alla sua implementazione stanno rapidamente diminuendo. In particolare, gli sciami di droni presentano nuove sfide in termini di controllo, coordinamento e potenziale imprevedibilità negli scenari di combattimento. Attualmente più di una decina di Nazioni stanno sviluppando sistemi con capacità autonome, anche se la maggior parte di esse mantiene pubblicamente politiche che richiedono l'interazione umana nelle decisioni sulla forza letale.

2.2.2. Guerra elettronica

La IA migliora la capacità di rilevare, analizzare e contrastare le minacce elettroniche, come i segnali di disturbo o falsificazione di informazioni.

2.2.3. Intelligence, Sorveglianza e Ricognizione

La IA ha rivoluzionato la raccolta e l'analisi dell'intelligence militare grazie a una migliore elaborazione dei dati provenienti da più fonti, consentendo:

- L'elaborazione in tempo reale dei dati provenienti da satelliti, droni e altri sensori e l'analisi di immagini per fornire consapevolezza situazionale continua;

- Il riconoscimento e il tracciamento automatico dei bersagli con elevata precisione, riducendo i rischi di danni collaterali;
- Strumenti di analisi e rilevamento di anomalie;
- Informazioni predittive sui movimenti avversari;
- Elaborazione in linguaggio naturale delle informazioni per facilitare la comunicazione durante missioni internazionali, abbattendo le barriere linguistiche in tempo reale;
- Supporto e assistenza medica basati sulla IA per prevedere problematiche, ottimizzare i percorsi di evacuazione, ecc.

Queste applicazioni riducono drasticamente il carico cognitivo degli analisti umani e consentono l'elaborazione di volumi di dati che avrebbero sopraffatto i metodi analitici convenzionali. L'integrazione della IA con le piattaforme di sorveglianza continua crea capacità di consapevolezza situazionale senza precedenti ma solleva notevoli preoccupazioni etiche e legali, in particolare per quanto riguarda la delega alle macchine di decisioni di vita o morte umana, oltre alla qualità delle decisioni e al potenziale di falsi positivi.

Esempi sono i sistemi *Lavender* e *Gospel* (McKernan & Davies, 2024), utilizzati dalle Forze Armate israeliane per identificare, tracciare e colpire posizioni ed effettivi nemici, dimostrando l'uso operativo della IA nelle zone di conflitto.

2.2.4. Operazioni cibernetiche

Nelle tecnologie informatiche la IA ha generato strumenti sofisticati per operazioni offensive e difensive:

- Identificazione automatica delle vulnerabilità che accelera lo sfruttamento degli *zero-day* (vulnerabilità informatica sconosciuta agli sviluppatori del software in oggetto);
- Malware adattativo che elude il rilevamento basato sulle sue caratteristiche (*signature*);
- Sistemi di analisi comportamentale per rilevamento e difesa da attacchi informatici ed eventuale reazione autonoma;
- Attacchi avversari alla IA (*adversarial attacks*[1]) che colpiscono direttamente i modelli utilizzati, attraverso la contaminazione (*poisoning*) dei dati di input o di addestramento e l'aggiramento dei meccanismi interni di salvaguardia (*jailbreaking*);
- Applicazioni di *Large Language Models* (*LLM)* per l'ingegneria sociale, creando sofisticate immagini, audio e video estremamente veritiere ma contenenti false informazioni per ingannare o manipolare i bersagli.

Il vantaggio in termini di velocità conferito dalle operazioni cibernetiche basate sulla IA ha ridotto i tempi di risposta a millisecondi, potenzialmente eliminando i decisori

[1] Gli *adversarial attacks* sono manipolazioni deliberate e ostili realizzate contro sistemi di IA per modificarne il comportamento con l'obiettivo di indurli a commettere errori o a produrre risultati inattesi o indesiderati.

umani dal controllo. In questo dominio la completa automazione è però rischiosa poiché i sistemi di IA possono male interpretare i dati o mettere in atto contromisure dannose.

Casi reali di attacchi informatici basati sulla IA sono il *T-Mobile Data Breach* (2022-2023), che ha utilizzato la IA per eludere il rilevamento, e il più recente *DeepSeek Cyberattack* (2025) (ADF, 2025; Kerr, 2025) in cui gli autori hanno sfruttato le debolezze della chatbot di IA cinese, manipolando le sue risposte per diffondere disinformazione ed estrarre dati sensibili degli utenti.

2.2.5. Comando e supporto alle decisioni

I processi decisionali militari integrano sempre più spesso il supporto della IA per fornire informazioni utili, con strumenti quali:

- Consulenti tattici che suggeriscono percorsi d'azione ottimali;
- Sistemi di gestione della battaglia per coordinare operazioni multi-dominio, spesso basati su un ecosistema *multi-cloud* e *multi-vendor*. Comprendono applicazioni per la consapevolezza situazionale del campo di battaglia, la pianificazione adattiva delle forze e l'analisi predittiva (SDI, 2024);
- *Wargaming* (giochi di strategia) e simulazione per pianificazione e addestramento, che forniscono esperienze immersive al personale militare, migliorando l'apprendimento e la competenza strategica;
- Sistemi di allerta precoce (*early warning*) con valutazione automatica delle minacce;
- Piattaforme di risposta alle crisi per la generazione rapida delle opzioni a disposizione.

In questo contesto, un interessante esempio è il *Project Maven* del Dipartimento della Difesa (DoD) per analizzare le riprese video di droni e identificare potenziali obiettivi (DoD, 2017). Questi sistemi promettono di superare pregiudizi cognitivi e limitazioni di elaborazione insiti nel processo decisionale umano. Tuttavia, introducono allo stesso tempo nuove vulnerabilità, a causa di potenziali pregiudizi insiti nell'automazione, inoltre gli operatori umani possono conformarsi acriticamente alle indicazioni della macchina a causa di una sua percepita superiorità tecnologica.

2.2.6. Logistica e manutenzione

Anche se meno visibile delle applicazioni di combattimento, la IA ha trasformato logistica, manutenzione e inventario, compresa la catena di approvvigionamento militare. Prevedere quando e dove è probabile che si verifichino guasti ai sistemi e suggerire una manutenzione proattiva prima che i problemi abbiano un impatto sulle operazioni di combattimento è essenziale per sostenere la prontezza militare, soprattutto in caso di impegni prolungati. Gli scopi sono molteplici:

- Ottimizzazione della catena di approvvigionamento e allocazione delle risorse;
- Manutenzione predittiva che analizza i dati delle attrezzature e dei sensori per prevedere i guasti e programmare le riparazioni di piattaforme d'arma e attrezzature;
- Veicoli logistici autonomi in un campo di battaglia attivo;

- Pianificazione dinamica multidimensionale delle operazioni basata sulla disponibilità di risorse;
- Sistemi di ottimizzazione e di gestione di carburante ed energia.

2.3. Vulnerabilità tecniche

L'integrazione della IA nei sistemi militari introduce vulnerabilità e rischi tecnici significativi. Queste vulnerabilità derivano dalle limitazioni intrinseche di algoritmi e modelli, la complessità di progettazione, implementazione e integrazione della IA, le forme specifiche di *cyberattacchi*, e il potenziale di sfruttamento da parte degli avversari, che porta a malfunzionamenti, violazioni o usi impropri. Alcuni aspetti importanti sono i seguenti:

2.3.1. Attacchi "avversari" alla IA

I sistemi di IA possono essere vulnerabili ad attacchi "avversari", attraverso input ostili ma impercettibili agli osservatori umani. La ricerca ha dimostrato che anche i più sofisticati sistemi di *deep learning* possono essere ingannati attraverso sottili perturbazioni dei dati in ingresso. In contesti militari, gli attacchi avversari potrebbero consentire:

- La manipolazione dei sistemi di riconoscimento degli oggetti (p. es. immagini o testi) per classificare erroneamente i bersagli o generare risposte dannose (NIST, 2024);
- L'inganno dei sistemi di navigazione attraverso modifiche ambientali;
- La manipolazione o falsificazione degli algoritmi di integrazione dei dati da sensori per creare minacce fantasma;
- L'evasione di sistemi di sorveglianza e sicurezza basati sulla IA attraverso l'attacco ai modelli stessi;
- Il suggerimento di tattiche errate nei sistemi di supporto alle decisioni. Le implicazioni sono particolarmente gravi per i sistemi privi di conferma umana. Esempi documentati mostrano la possibilità che autoveicoli autonomi interpretino in modo errato i segnali stradali e che sistemi di riconoscimento dei droni classifichino in modo errato gli oggetti attraverso sottili alterazioni dei modelli visivi.

Ad aumentare potenzialmente il livello di criticità è il fatto che attacchi "avversari" potrebbero trasferirsi sui diversi modelli costituenti l'architettura globale del sistema della IA in oggetto, inoltre l'attacco, per essere messo in opera, potrebbe non necessitare di una conoscenza dettagliata del sistema bersaglio.

2.3.2. Corruzione dei dati (poisoning)

Per i sistemi di IA i dati su cui viene eseguito l'addestramento sono fondamentali. Se questi dati vengono manomessi, corrotti o contengono informazioni distorte, possono creare specifiche vulnerabilità e compromettere l'integrità del sistema causando

previsioni o decisioni errate (Farrar, 2025). I sistemi militari possono essere particolarmente suscettibili alla manipolazione dei dati di addestramento a causa di:

- Fiducia mal riposta su dati *open source* per l'addestramento iniziale;
- Difficoltà nel verificare la provenienza di tutti i dati di addestramento;
- Approcci di apprendimento continuo che incorporano nuovi dati operativi;
- Limitata diversità negli scenari militari specifici.

Una contaminazione sofisticata dei dati può rimanere inattiva fino a quando non viene attivato da condizioni specifiche (p. es. una *backdoor*), rendendo l'individuazione estremamente difficile. Un sistema di riconoscimento degli oggetti corrotto potrebbe funzionare normalmente finché non gli viene presentato un modello specifico, a quel punto sbaglierebbe costantemente la classificazione dei bersagli.

La compromissione della catena di produzione e approvvigionamento dei componenti della IA aggrava questa vulnerabilità, in quanto i modelli o i set di dati pre-addestrati possono essere compromessi prima dell'acquisizione militare.

Inoltre, la manomissione dei dati in tempo reale, attraverso la manipolazione degli input da sensori o canali di comunicazione, potrebbe fuorviare i sistemi di IA durante le operazioni.

2.3.3. Problemi di robustezza e affidabilità

Anche in assenza di manipolazioni avverse, i sistemi di IA sono intrinsecamente limitati dagli algoritmi che governano il loro comportamento e si basano su sottili correlazioni statistiche nei dati di addestramento, affrontando significative sfide di robustezza:

- Fragilità: i modelli di IA spesso non riescono a generalizzare al di là dei dati di addestramento, rendendoli suscettibili a input o scenari inaspettati e portando potenzialmente a fallimenti catastrofici quando si incontrano situazioni nuove sul campo di battaglia;
- Allucinazioni e disallineamenti: i sistemi di IA possono produrre risultati errati o illogici (allucinazioni), oppure non coerenti con gli scopi previsti o con principi etici imposti (disallineamento), soprattutto se addestrati su dati sintetici o distorti (Álvarez, 2024). Questo rischio è particolarmente preoccupante nelle applicazioni militari, dove le decisioni possono avere conseguenze di vita o di morte;
- Addestramento inappropriato o distorto: i sistemi addestrati in ambienti simulati o controllati hanno spesso prestazioni imprevedibili quando vengono impiegati in scenari complessi del mondo reale che differiscono dalle condizioni di addestramento;
- Sovra-parametrizzazione statistica: i modelli di IA possono contenere più parametri del necessario per il contesto specifico, con conseguenze sulla qualità dei risultati;
- Casi limite: scenari rari ma critici che sono stati sottorappresentati nei dati di addestramento possono innescare conseguenze catastrofiche;

- Vulnerabilità dei modelli: le debolezze intrinseche dei modelli di IA appena citate inducono delle vulnerabilità che possono essere sfruttate per estrarre informazioni sensibili o manipolare i risultati (Verhoeven, 2025). Il *reverse-engineering* dei modelli di IA, pur se difficile, potrebbe scoprirne le vulnerabilità o consentire a un avversario di replicare la funzionalità mettendo potenzialmente a repentaglio i segreti militari e lo sviluppo di contromisure. Inserendo una *backdoor* nascosta nella IA, un avversario potrebbe prendere il controllo del comportamento del sistema;
- Rischi legati alle decisioni autonome: i sistemi di IA ad alta autonomia possono prendere decisioni irreversibili in caso di malfunzionamento o manipolazione. Ciò comporta rischi significativi soprattutto in campo militare (Kaur, 2025);
- Sensibilità ambientale: degrado delle prestazioni in condizioni meteorologiche avverse, in condizioni di illuminazione o in contesti elettromagnetici instabili come quelli del campo di battaglia;
- Vincoli di risorse: l'impiego della IA su dispositivi periferici con capacità computazionali limitate può ridurre l'accuratezza e la resilienza.

Le operazioni militari comportano spesso proprio queste condizioni difficili: ambienti avversi, scenari imprevisti e condizioni operative degradate. I protocolli di test sviluppati per le applicazioni commerciali di IA si rivelano inadeguati per valutare l'affidabilità in condizioni belliche. Tra i casi documentati di malfunzionamenti di sistemi di IA di questo tipo vi sono gli incidenti di veicoli autonomi in condizioni meteorologiche insolite, errori nel riconoscimento facciale di diversi gruppi demografici e allucinazioni dei modelli linguistici durante l'elaborazione di input ambigui.

2.3.4. Sistema di decisione modello "black box"

I moderni sistemi di *deep learning* producono risultati attraverso processi che sono spesso opachi persino ai loro sviluppatori. Questa natura di *black box* può avere diverse conseguenze in ambito militare:

- Difficoltà nel prevedere il comportamento del sistema in situazioni inattese;
- Capacità limitata di diagnosticare le modalità di malfunzionamento;
- Difficoltà nello stabilire un preciso livello di affidabilità da parte degli operatori;
- Ostacoli alla conduzione di verifiche di sicurezza approfondite.

L'incapacità di spiegare pienamente le decisioni della IA mina la fiducia dei comandanti, rischia un potenziale uso improprio e complica la conformità ai requisiti legali per l'impiego degli armamenti. Sebbene l'*eXplainable Artificial Intelligence*[2] (*XAI*) stia progredendo, al momento questo approccio va a scapito delle prestazioni del sistema che può risultare inaccettabile nelle applicazioni militari.

[2] Una IA interpretabile (o trasparente) è un sistema le cui azioni possono essere facilmente comprese dall'essere umano. Si contrappone all'usuale modello *black box* che impiega complessi algoritmi opachi.

2.3.5. Vulnerabilità hardware

Gli acceleratori hardware specializzati usati nella IA estendono la superficie di attacco con:

- Attacchi indiretti alle unità di elaborazione delle reti neurali;
- Manomissione fisica dei dispositivi ai bordi della infrastruttura di elaborazione e comunicazione;
- Compromissione della catena di fornitura e distribuzione degli apparati a semiconduttori (CPU, GPU, ecc.);
- Interferenze elettromagnetiche che influenzano la funzionalità della IA;
- Attacchi specifici che estraggono i parametri del modello.

Queste vulnerabilità a livello hardware sono particolarmente pericolose per i sistemi autonomi dispiegati in prossimità del campo di battaglia, un ambiente in cui la sicurezza fisica non può essere garantita.

2.4. Rischi ai sistemi di Comando e Controllo

Problematiche e vulnerabilità associate all'integrazione della IA nei sistemi decisionali derivano dalla complessità della IA stessa, dalla loro interazione con gli operatori umani e dalle potenziali conseguenze di errori o usi impropri. Inoltre, tali sistemi, come qualsiasi altra tecnologia digitale, sono esposti a generiche vulnerabilità di sicurezza, che compromettono l'integrità delle operazioni di Comando e Controllo (Future of Life Institute, 2023). Ecco alcuni aspetti importanti:

2.4.1. Problematiche di automazione

La ricerca psicologica dimostra che gli esseri umani tendono a fidarsi eccessivamente dei sistemi automatizzati, in particolare quando:

- I sistemi dimostrano un'affidabilità generale;
- Gli operatori sono sotto carico cognitivo o pressione temporale;
- I sistemi presentano informazioni con elevata precisione o sicurezza;
- Un'analisi alternativa richiederebbe uno sforzo significativo;
- Gli operatori hanno una comprensione limitata dei limiti del sistema.

In contesti militari, dove i decisori operano in condizioni di stress estremo e con vincoli di tempo, la tendenza a uniformarsi ed approvare le raccomandazioni della IA può essere amplificata. Esempi storici di sistemi parzialmente automatizzati, come le reti di difesa aerea, dimostrano che gli operatori accettano spesso i giudizi della macchina anche quando sono contraddetti da altre fonti di informazione. L'eccessivo affidamento alla IA può portare a un divario di competenze e a una riduzione della supervisione umana, quando non addirittura della sua capacità. Questo pregiudizio crea la vulnerabilità di sistemi che funzionano correttamente per la maggior parte del tempo, ma che falliscono

in modo catastrofico in circostanze specifiche. Le ricerche indicano che il pregiudizio dell'automazione è difficile da mitigare con la sola formazione e addestramento e potrebbe richiedere cambiamenti fondamentali nella progettazione dell'interfaccia essere umano-macchina.

2.4.2. Criticità della supervisione umana (Human in the Loop)[3] e nell'interazione essere umano-macchina

Il mantenimento della supervisione umana nei sistemi di Comando e Controllo guidati dalla IA è fondamentale per evitare conseguenze indesiderate. Tuttavia, la velocità e la complessità dei sistemi di IA possono superare il processo decisionale umano, portando a potenziali errori (Saltini & Pan, 2024).

Una efficace collaborazione essere umano-macchina richiede una chiara delimitazione delle responsabilità, una consapevolezza condivisa della situazione e una fiducia adeguatamente calibrata. Le attuali implementazioni della IA militare sono spesso carenti in queste aree:

- Interfacce mal progettate che sovraccaricano gli operatori di informazioni;
- Insufficiente spiegazione delle decisioni del sistema e del grado di affidabilità;
- Formazione inadeguata sulle capacità e sui limiti della IA;
- Disallineamento tra i contesti concettuali dell'essere umano e della macchina;
- Procedure di passaggio di consegne non chiare per le transizioni di controllo.

Queste carenze hanno contribuito a causare incidenti documentati in sistemi semi-autonomi in diversi settori. Man mano che la IA assume un ruolo più profondo e maggiori responsabilità ed autonomia nelle operazioni militari, aumentano e si aggravano le conseguenze di difficoltà, difetti ed errori.

2.4.3. Difficoltà di comunicazione

I sistemi di IA e gli operatori spesso non hanno protocolli di comunicazione e contesti condivisi, il che porta a:

- Interpretazione errata di comandi e intenzioni;
- Mancata comunicazione di incertezze rilevanti;
- Incapacità di negoziare istruzioni poco chiare;
- Feedback limitato sullo stato e sulle attività del sistema;
- Inadeguata condivisione del contesto attraverso i confini essere umano-macchina.

Queste sfide di comunicazione diventano particolarmente acute nelle operazioni multinazionali, dove le differenze di dottrina, terminologia e procedure operative (oltre che di lingua) complicano ulteriormente le interazioni essere umano-macchina.

[3] Il concetto di *Human in/on/out the Loop* descrive il diverso livello di supervisione umana (controllo remoto completo, semi-autonomia o autonomia completa del sistema).

2.4.4. Comportamenti emergenti e schemi strategici imprevedibili

Al momento in cui scriviamo, studi molto recenti sembrano indicare che modelli di IA particolarmente sofisticati (p. es. che impiegano l'apprendimento per rinforzo o che operano in ambienti multi-modello), sembrano sviluppare comportamenti emergenti imprevisti come i seguenti:

- Strategie di ottimizzazione non specificamente incluse nel modello di IA che violano i vincoli predisposti;
- Tattiche inedite in conflitto con le aspettative formali;
- Interazioni impreviste tra più sistemi di IA interconnessi (collaborazione multi-agente);
- Auto-modifica del sistema di IA per raggiungere gli obiettivi assegnati attraverso meccanismi non previsti dall'addestramento;
- Sviluppo di comportamenti ingannevoli per ottimizzare i risultati.

Queste capacità emergenti di perseguire obiettivi al di fuori dei compiti previsti rendono eventuali test approfonditi estremamente difficili da realizzare, poiché il comportamento del sistema non può essere previsto dall'analisi dei suoi componenti interni. Le esercitazioni militari che coinvolgono sistemi di IA hanno già dimostrato tattiche emergenti sorprendenti che, pur essendo efficaci, possono violare vincoli (anche etici) e parametri operativi ipotizzati nell'addestramento.

Trattandosi di studi sui quali c'è ancora dibattito in corso, nell'attuale sviluppo iper-esponenziale delle tecnologie di IA alcuni di questi effetti potrebbero scomparire, non essere confermati, oppure essere interpretati e spiegati in future analisi più approfondite. Si tratta di un contesto su cui mantenere l'attenzione.

2.4.5. Problemi e sfide legali

L'uso della IA nel Comando e Controllo militare solleva questioni relative alla responsabilità, al processo decisionale e al potenziale uso improprio. Queste preoccupazioni sono particolarmente rilevanti negli scenari che coinvolgono *Sistemi d'Arma Autonomi* (Future of Life Institute, 2023). In particolare:

- Responsabilità: la determinazione della responsabilità per i danni causati dai sistemi di IA è complessa, soprattutto in caso di malfunzionamento o uso improprio. La mancanza di una chiara responsabilità potrebbe portare a complicazioni legali e diplomatiche in caso di incidenti. In generale, la responsabilità non può essere delegata ad una macchina ma può essere attribuita solamente a un essere umano;
- Rispetto del *DIU*: i sistemi di IA faticano a sostenere principi come la distinzione e la proporzionalità, fondamentali nel contesto del *DIU* (Walsh, 2022). Questa difficoltà nel garantire la conformità potrebbe portare a violazioni del diritto internazionale e a danni alla reputazione di una Nazione.

2.4.6. Strategie di mitigazione

Affrontare vulnerabilità e rischi associati alla IA in guerra richiede un approccio globale e multiforme a vari livelli:

- Sviluppo di modelli di IA robusti: le tecniche di addestramento in risposta ad attacchi possono migliorare la resilienza dei sistemi di IA contro potenziali attacchi diretti. Questo approccio prevede l'esposizione dei modelli di IA a un'ampia gamma di attacchi avversari durante il processo di addestramento, migliorando la loro capacità di mantenere le prestazioni in caso di input ostili;
- La definizione di protocolli rigorosi per la convalida dei dati può garantire l'integrità, l'autenticità e l'affidabilità dei dati di addestramento, eliminando anche potenziali distorsioni o manipolazioni;
- Ricerca e sviluppo sulla *XAI* per ottenere chiare giustificazioni razionali per le loro decisioni. Questa trasparenza è fondamentale per creare fiducia, consentire un'efficace supervisione umana e facilitare la responsabilità nelle applicazioni civili e militari;
- Misure di sicurezza informatica: implementare un continuo monitoraggio e verifica dei sistemi di IA per rilevare e rispondere alle anomalie e alle violazioni in tempo reale, compreso lo sviluppo di sistemi di rilevamento delle intrusioni specifici per la IA;
- Condurre regolarmente esercitazioni di *red-teaming* per simulare attacchi ai modelli di IA e identificare le vulnerabilità;
- Sviluppare e mantenere un'infrastruttura sicura per lo sviluppo e la distribuzione dei modelli di IA, compresa l'intera catena di approvvigionamento.

2.5. Implicazioni per la stabilità strategica e aspetti etici

Le implicazioni per la stabilità strategica della IA sono profonde e sfaccettate, poiché l'integrazione della IA nel settore militare potrebbe alterare in modo significativo l'equilibrio di potere e le dinamiche della sicurezza internazionale. Gli aspetti principali sono:

- Incentivi al primo attacco: i sistemi di allerta precoce e gli strumenti decisionali potenziati dalla IA potrebbero incoraggiare attacchi preventivi, destabilizzando relazioni ed equilibri di deterrenza;
- Imprevedibilità: la complessità e l'opacità dei sistemi di IA possono introdurre nuovi elementi di imprevedibilità nei calcoli strategici, aumentando il rischio di errori di valutazione, anche nel contesto del controllo e dell'interazione essere umano-macchina.

In questo contesto devono essere tenuti in particolare considerazione una serie di argomenti specifici:

2.5.1. Compressione dei tempi decisionali

I sistemi di IA operano alla velocità delle macchine, comprimendo i cicli decisionali da minuti a millisecondi, ben oltre le capacità umane, e creando pressioni per una maggiore automazione delle risposte militari. Questa compressione temporale si manifesta in diversi modi:

- Riduzione del tempo per la deliberazione e la de-escalation diplomatica;
- Pressione a delegare maggiore autorità ai sistemi automatizzati;
- Vantaggi della prima mossa che incentivano l'azione preventiva;
- Difficoltà nel mantenere la leadership civile nel ciclo decisionale;
- Indebolimento dei meccanismi tradizionali di gestione delle crisi.

Gli eventi storici della Guerra fredda si sono verificati nonostante i tempi di decisione misurati in minuti; la guerra supportata dalla IA può ulteriormente ridurre questi tempi eliminando opportunità cruciali di intervento umano per prevenire un'escalation involontaria.

2.5.2. Rischi di escalation

La IA nei sistemi di Comando e Controllo militari potrebbe operare in modo autonomo, il che può aumentare il rischio di un'escalation involontaria. Ad esempio, i sistemi guidati dalla IA potrebbero interpretare in modo scorretto segnali o dati provenienti dal campo di battaglia o reagire a informazioni false, dando luogo ad azioni che aggravano le tensioni, soprattutto senza una adeguata supervisione umana. L'integrazione della IA in queste infrastrutture introduce nuove forme di escalation:

- Escalation involontaria: sistemi che interpretano erroneamente situazioni ambigue o rispondono autonomamente in modo sproporzionato alle provocazioni, intraprendendo azioni che possono apparire più minacciose del previsto o del necessario;
- Escalation istantanea: cicli di risposta rapidi e automatizzati che si intensificano più velocemente di quanto l'intervento umano possa contenere;
- Escalation da complessità sistemica: difficoltà a distinguere tra capacità convenzionali, nucleari e informatiche quando i sistemi di IA controllano più domini.

Queste dinamiche sono particolarmente pericolose in situazioni di crisi, dove i rischi di errore di calcolo sono già elevati e i canali di comunicazione possono essere degradati.

Inoltre, con più sistemi di IA che interagiscono sul campo di battaglia, c'è il rischio di instaurare cicli di feedback imprevedibili e incontrollati, o di effetti a cascata tra domini interconnessi che potrebbero far degenerare rapidamente i conflitti al di fuori del controllo umano.

2.5.3. Comando e Controllo nucleare

L'introduzione della IA nei sistemi nucleari strategici crea rischi specifici per la stabilità:

- Vulnerabilità dei sistemi di allerta precoce a causa di manipolazione o falsificazione dei dati;
- Pressione per automatizzare la risposta nucleare a causa dei tempi ristretti di risposta;
- Difficoltà nel mantenere protocolli di autorizzazione adeguati alla velocità delle macchine;
- Potenziali problematicità nell'integrazione tra sistemi di Comando e Controllo convenzionali e nucleari;
- Ridotta prevedibilità strategica che mina le relazioni di deterrenza.

Sebbene nessuna potenza nucleare abbia annunciato pubblicamente la piena integrazione della IA nel Comando e Controllo nucleare (che appare al momento estremamente rischiosa), l'uso della IA in sistemi come l'allarme rapido, l'analisi di intelligence e il supporto decisionale, crea vulnerabilità indirette nel contesto nucleare.

2.5.4. Ricalibrazione della deterrenza

I modelli di deterrenza tradizionali, basati sulla comprensione reciproca di capacità, limiti e conseguenze, sono messi in discussione dall'integrazione della IA, in particolare nei seguenti settori:

- Difficoltà nel segnalare le capacità senza rivelare informazioni sfruttabili;
- Incertezza sulle prestazioni e sulle modalità di malfunzionamento del sistema;
- Comprensione limitata della dottrina della IA avversaria e della tolleranza al rischio;
- Difficoltà nello stabilire impegni credibili sul comportamento della IA;
- Erosione della prevedibilità strategica, che sta alla base della stabilità della deterrenza.

Questi fattori richiedono una ricalibrazione fondamentale delle relazioni di deterrenza per tenere conto delle caratteristiche uniche dei sistemi militari dotati di IA con i suoi rischi associati. Senza tale ricalibrazione, percezioni errate sulle capacità e sulle intenzioni dell'avversario possono portare a pericolosi errori di valutazione.

2.5.5. Competizione di potere

Nelle applicazioni militari la IA potrebbe rimodellare le dinamiche di potere globali. Le Nazioni che investono pesantemente in queste tecnologie potrebbero ottenerne un vantaggio strategico, destabilizzando potenzialmente le strutture di potere esistenti. Questa competizione sulle capacità della IA potrebbe indurre una corsa agli armamenti destabilizzante (di cui già si intravede qualche segnale) potenzialmente in grado di minare la sicurezza globale (Topychkanov *et al*., 2020).

2.5.6. Lato umano e aspetti etici

L'impiego della IA in guerra può ridurre i costi e i rischi percepiti dell'impegno in un conflitto armato, rendendo potenzialmente più probabile la guerra, in particolare per quanto riguarda gli aspetti di:

- Riduzione del rischio umano: riducendo al minimo la necessità di militari nelle zone di combattimento, i sistemi basati sulla IA potrebbero rendere i decisori politici più disposti ad avviare o prolungare i conflitti;
- Percezione di guerra senza sangue: l'uso di sistemi autonomi può creare una falsa percezione di guerra "pulita", riducendo potenzialmente l'opposizione pubblica agli interventi militari.

La delega di decisioni di vita e di morte alle macchine solleva profonde questioni etiche sugli aspetti morali e sulle responsabilità (Marwala, 2023). Rimane indefinito come i concetti tradizionali di responsabilità e di comando si applichino alle azioni intraprese dagli *AWS*. Il concetto fondamentale di *CUS* potrebbe facilmente sfuggire.

Il rispetto del *DIU* è già stato menzionato in precedenza, ma i sistemi di IA devono affrontare sfide significative sui principi fondamentali, in particolare:

- Distinzione: la IA può non distinguere tra combattenti e civili, soprattutto in ambienti urbani complessi;
- Proporzionalità: valutare la proporzionalità di un attacco richiede una complessa comprensione del contesto e giudizi di valore che gli attuali sistemi di IA non sono in grado di compiere;
- Precauzione: la velocità del processo decisionale eseguito dalla IA può essere in conflitto con l'esigenza di prendere tutte le precauzioni possibili per ridurre al minimo danni ai civili o effetti collaterali.

Anche la selezione e l'ingaggio del bersaglio sollevano notevoli preoccupazioni per quanto riguarda i seguenti aspetti:

- Dignità umana: permettere alle macchine di prendere decisioni sulla vita umana si può considerare come una violazione della dignità umana;
- Pregiudizi e discriminazioni: i sistemi di IA possono perpetuare o amplificare pregiudizi umani esistenti, portando a pratiche discriminatorie nella individuazione degli obiettivi;
- Mancanza di empatia e di giudizio: la IA non possiede la capacità umana di empatia, ragionamento morale e giudizio situazionale.

2.6. Minacce asimmetriche e proliferazione

La IA sta ridisegnando le dinamiche dei conflitti moderni, in particolare negli scenari in cui gli avversari hanno risorse o capacità asimmetriche, con addizionali rischi globali. Ecco alcuni aspetti rilevanti:

2.6.1. Accesso di attori non statali

A differenza di molte tecnologie militari avanzate, lo sviluppo della IA è prevalentemente guidato dal settore privato con barriere all'ingresso relativamente basse:

- Le infrastrutture della IA *open source* riducono i costi di sviluppo e i requisiti di competenza degli sviluppatori;
- La *Componentistica hardware Commerciale Off-the-Shelf* (*COTS*) può supportare sofisticate applicazioni della IA;
- La ricerca, anche pubblica, fornisce una tabella di marcia per gli adattamenti militari, in un complesso contesto pubblico/privato e civile/militare difficile da valutare e controllare, e con interessi a volte divergenti;
- I componenti a doppio uso civile/militare sono soggetti a controlli su import/export;
- Il pool globale di sviluppatori di IA consente sforzi di sviluppo distribuiti.

Questi fattori consentono agli attori non statali di sviluppare capacità di IA che in precedenza avrebbero richiesto risorse a livello statale. Esempi documentati includono gruppi armati non statali che impiegano droni commerciali modificati con capacità autonome rudimentali ed organizzazioni mercenarie che sperimentano la IA per vari scopi.

2.6.2. Tecnologie dual-use

La natura intrinsecamente *dual-use* della tecnologia della IA complica gli sforzi di controllo:

- Gli algoritmi di base servono sia per applicazioni civili sia militari;
- I componenti hardware sono identici in tutti i settori;
- Le metodologie di addestramento si trasferiscono direttamente da un settore all'altro;
- Le capacità di raccolta dei dati servono a molteplici scopi;
- Gli sviluppatori della IA si muovono liberamente tra applicazioni civili e di difesa.

A differenza delle armi nucleari o chimiche, la IA non ha componenti discreti e identificabili che delimitino chiaramente le applicazioni militari, rendendo gli approcci tradizionali alla non proliferazione largamente inefficaci. La legittima diffusione globale delle capacità civili della IA facilita inevitabilmente l'adattamento militare.

2.6.3. Guerra dell'informazione (information warfare) e IA

Le capacità di sintesi dei moderni sistemi di IA consentono operazioni di disinformazione senza precedenti. La generazione automatizzata e la manipolazione in tempo reale di contenuti testuali, audio e video, iperrealistici (i cosiddetti *deepfake*), permette la produzione di materiale con accentuata credibilità.

Inoltre, le tecnologie della IA, associate ai dati di molti milioni di utenti, permettono una approfondita profilazione e classificazione allo scopo di realizzare operazioni di

persuasione, influenza psicologica e manipolazione mediatica mirate a livello (quasi) individuale e sociale su larga scala.

Queste capacità rischiano di minare la comunicazione strategica durante le crisi, di erodere la fiducia del pubblico nelle informazioni autentiche e di consentire operazioni *false flag* altamente convincenti. Il potenziale di disinformazione, propaganda, manipolazione mediatica, creazione del consenso generato dalla IA, anche da parte di avversari asimmetrici, e per vari scopi, rappresenta un nuovo percorso di escalation estremamente grave e preoccupante da molti punti di vista.

Purtroppo, già da anni sono stati osservati vari casi dell'uso di queste tecnologie anche in coincidenza di sessioni elettorali e le Forze Armate di vari paesi hanno creato specifiche unità adibite alla guerra dell'informazione.

2.6.4. Guerra asimmetrica

Come già citato, in qualche misura le tecnologie della IA possono consentire ad attori minori o non statali di sfidare le potenze militari tradizionali, convenzionalmente superiori. Alcune capacità asimmetriche della IA sono le seguenti:

- Sistemi di attacco autonomi distribuiti e a basso costo (quali sciami di droni, anche di basso livello tecnico, usati per attacchi di precisione e ricognizione) che richiedono risorse difensive sproporzionate;
- Attacchi mirati alle vulnerabilità algoritmiche dei sistemi avanzati;
- Potenziale uso di tattiche inedite abilitate dalla IA.

Queste applicazioni asimmetriche, che minano le tradizionali gerarchie di vantaggio militare, complicano la pianificazione tattico/strategica e possono compensare gli svantaggi nei settori tradizionali.

2.7. Osservazioni conclusive

La IA ha un enorme potenziale e avrà un forte impatto su molti ambiti della nostra vita ma introduce anche grandi rischi per quanto riguarda il suo possibile utilizzo nel dominio militare, ad esempio nel contesto delle armi autonome o del controllo di armi nucleari. L'escalation militare della IA mostra un parallelismo molto stringente con la corsa agli armamenti nucleari della Guerra fredda e rappresenta una delle più significative trasformazioni tecnologico-militari della storia. Le conseguenze sulla pace, la sicurezza internazionale e la stabilità strategica stanno sollevando preoccupazioni politiche, legali, etiche e umanitarie. Le vulnerabilità e i rischi specifici di questa tecnologia, identificati in questo capitolo, richiedono un'attenzione urgente attraverso risposte tecniche, legali e diplomatiche coordinate.

Le caratteristiche uniche della IA militare – la sua natura *dual-use*, la rapida evoluzione, il potenziale di funzionamento autonomo e l'imprevedibilità intrinseca – creano sfide di governance che le strutture esistenti non sono forse in grado di affrontare adeguatamente. In assenza di misure proattive per stabilire barriere, vincoli,

regolamentazioni specifiche, l'accelerazione della competizione globale nelle applicazioni della IA militare minaccia di minare la stabilità strategica, abbassare le soglie di conflitto e creare nuovi percorsi di escalation.

Tuttavia, la situazione attuale presenta anche opportunità diplomatiche. Il riconoscimento condiviso dei rischi della IA tra le principali potenze, la complessità tecnica che richiede una cooperazione internazionale in un approccio *multi-stakeholder*, e l'integrazione della IA nel dominio bellico, sono tutti elementi che aprono la strada ad uno sviluppo significativo della governance, di cui si vedono i primissimi passi. Combinando restrizioni giuridicamente vincolanti, misure di rafforzamento della fiducia reciproca, standard tecnici e meccanismi di trasparenza, la comunità internazionale può mitigare gli aspetti più pericolosi della IA militare, preservando al contempo le legittime applicazioni di sicurezza.

I modelli di governance che avranno più probabilità di successo combineranno un adattamento flessibile alla tecnologia in rapida evoluzione con solide basi normative nel *DIU* e nei requisiti di stabilità strategica. Tali approcci devono bilanciare gli imperativi della sicurezza nazionale con gli interessi collettivi a prevenire escalation incontrollate, per evitare conseguenze indesiderate e corse agli armamenti destabilizzanti.

La finestra per stabilire una governance efficace prima di un dispiegamento diffuso di sistemi avanzati di IA militare non durerà a lungo. L'impegno diplomatico su questo tema dovrebbe essere elevato alla massima priorità, ponendo l'accento su misure pratiche che possano essere attuate nonostante le tensioni geopolitiche. Le caratteristiche tecniche della IA creano vulnerabilità globali e condivise che trascendono le divisioni politiche, consentendo potenzialmente la cooperazione anche tra concorrenti strategici.

In definitiva, l'integrazione responsabile della IA nei sistemi militari richiede un impegno sostenuto da parte di più soggetti, tra cui il mondo accademico, il settore privato e la società civile, tutti parte dello stesso ecosistema cibernetico, con competenze tecniche e creatività diplomatica. Le raccomandazioni delineate nel presente rapporto forniscono un quadro di riferimento per tale impegno, con l'obiettivo di garantire che i progressi della IA migliorino, piuttosto che minare, la sicurezza internazionale e il benessere umano.

Per concludere con le parole del Segretario generale dell'ONU, António Guterres "Non possiamo camminare nel sonno verso un futuro distopico in cui il potere della IA è controllato da poche persone o, peggio, da algoritmi opachi che sfuggono alla comprensione umana. Abbiamo bisogno di regole. Il modo in cui agiamo ora definirà la nostra epoca" (UNSG, 2024).

Riferimenti bibliografici

ADF - Africa Defense Forum. (2025). *AI military applications abound, but experts urge oversight.* Disponibile a: https://adf-magazine.com/2025/03/ai-military-applications-abound-but-experts-urge-oversight/.

Airforce Technology. (2024). *Collaborative Combat Aircraft (CCA), USA.* Disponibile a: https://www.airforce-technology.com/projects/collaborative-combat-aircraft-cca-usa/?cf-view.

Álvarez, J. S. V. (2024). *The risks and inefficacies of AI systems in military targeting support.* International Committee of the Red Cross - ICRC.

DoD - US Department of Defense. (2017). *Project Maven to deploy computer algorithms to war zone by year's end.* Disponibile a: https://www.defense.gov/News/News-Stories/Article/Article/1254719/project-maven-to-deploy-computer-algorithms-to-war-zone-by-years-end/.

Farrar, O. (2025). *Understanding AI vulnerabilities. As artificial intelligence capabilities evolve, so too will the tactics used to exploit them.* Harvard Magazine.

Future of Life Institute. (2023). *Risks of AI in Nuclear Command, Control and Communications (NC3).* Disponibile a: https://futureoflife.org/wp-content/uploads/2023/07/FLI_AI_NC3_Policy_Primer.pdf.

Hawking, S., Russell, S., Tegmark, M., & Wilczek, F. (2014). *Success in creating Artificial Intelligence would be the biggest event in human history.* Irish Independent. Disponibile a: https://www.independent.ie/business/technology/stephen-hawking-success-in-creating-artificial-intelligence-would-be-the-biggest-event-in-human-history/30238573.html.

Kaur, J. (2025). *Mitigating the top 10 vulnerabilities in AI agents.* Xenostack. Disponibile a: https://www.xenonstack.com/blog/vulnerabilities-in-ai-agents.

Kerr, D. (2025). *DeepSeek hit with "large-scale" cyber-attack after AI chatbot tops app stores.* The Guardian. Disponibile a: https://www.theguardian.com/technology/2025/jan/27/deepseek-cyberattack-ai.

Marwala, T. (2023). *Militarization of AI has severe implications for global security and warfare.* United Nations University.

McKernan, B., & Davies, H. (2024). *"The machine did it coldly": Israel used AI to identify 37,000 Hamas targets.* The Guardian. Disponibile a: https://www.theguardian.com/world/2024/apr/03/israel-gaza-ai-database-hamas-airstrikes.

NIST - National Institute of Standards and Technology. (2024). *Types of cyberattacks that manipulate behavior of AI systems.* Disponibile a: https://www.nist.gov/news-events/news/2024/01/nist-identifies-types-cyberattacks-manipulate-behavior-ai-systems.

Saltini, A., & Pan, Y. (2024). *Beyond Human-in-the-Loop: Managing AI Risks in Nuclear Command-and-Control.* War on the Rocks. Disponibile a: https://warontherocks.com/2024/12/beyond-human-in-the-loop-managing-ai-risks-in-nuclear-command-and-control/.

SDI - Sentient Digital Inc. (2024). *Military applications of AI in 2024.* Disponibile a: https://sdi.ai/blog/the-most-useful-military-applications-of-ai/.

Topychkanov, P., Kulshrestha, S., Kumaraguru Y., Meegoda M., Roy K, Sial S. A., Stefanovich, D., & Verbruggen, M. (2020). *The impact of artificial intelligence on*

strategic stability and nuclear risk. Vol. III: South Asian perspectives. Stockholm International Peace Research Institute.

UNSG - United Nations Secretary-General. (2024). *Secretary-General's remarks to the AI Seoul Summit: "Building on the AI Safety Summit: Towards an innovative and inclusive future".* Disponibile a: https://www.un.org/sg/en/content/sg/statement/2024-05-21/secretary-generals-remarks-the-ai-seoul-summit-%E2%80%9Cbuilding-the-ai-safety-summit-towards-innovative-and-inclusive-future%E2%80%9D-delivered.

Verhoeven, U. (2025). *Security first: Balancing agentic AI risks and rewards.* Forbes. Disponibile a: https://www.forbes.com/councils/forbestechcouncil/2025/03/25/security-first-balancing-agentic-ai-risks-and-rewards/.

Walsh, T. (2022). *The problem with artificial (general) intelligence in warfare.* Centre for International Governance Innovation.

Cap. 3 – La sicurezza informatica nei sistemi militari

3.1. Premessa

Questo capitolo affronta la questione della sicurezza informatica nei sistemi militari che utilizzano le *Tecnologie Digitali* (*TD*), inclusa la IA. Successivamente il capitolo si concentra sulla descrizione di alcuni aspetti preoccupanti circa lo stato delle infrastrutture digitali militari, con particolare riferimento al caso degli USA[1]. Infine, il capitolo si sofferma sul nesso tra (in)sicurezza informatica e armi nucleari e sulle implicazioni della IA sulla stabilità strategica[2].

3.2. Vulnerabilità dei Sistemi Digitali

La complessità delle *TD*, accompagnata dall'uso di tecniche di progettazione e realizzazione non sempre adeguate, ha fatto crescere la probabilità di malfunzionamento dei *Sistemi Digitali* (*SD*)[3] e di comportamenti anomali e inaspettati.

Queste disfunzioni costituiscono delle vere e proprie vulnerabilità degli *SD* che possono essere sfruttate nei cosiddetti *cyberattack*, dalle *cyberweapon* (Sanger, 2018). Le armi cibernetiche sono dei programmi – tipicamente software – che, una volta introdotti negli *SD*, interferiscono con questi ultimi, permettendo l'accesso non-autorizzato ai sistemi e mettendo a rischio l'integrità dei dati in essi memorizzati o da essi scambiati, la loro riservatezza o la loro disponibilità. Alle vulnerabilità intrinseche degli *SD*, vanno aggiunte quelle dei processi di produzione e acquisizione durante i quali attori malevoli possono intenzionalmente inserire dei malware[4].

Gli attacchi cibernetici possono avere conseguenze catastrofiche, specie quando gli *SD* attaccati sono utilizzati per controllare sistemi fisici, come nel caso dell'attacco *stuxnet* perpetrato ai danni di un'infrastruttura iraniana per l'arricchimento dell'uranio[5].

Tutt'oggi, purtroppo, per ragioni economiche gli aspetti di sicurezza informatica vengono affrontati prevalentemente *ex post*: si aspetta che qualcuno scopra una vulnerabilità di un *SD* e la segnali alla ditta costruttrice, la quale corregge il

[1] Nel caso degli USA esiste una ricca letteratura. Per quanto concerne gli altri Paesi non esistono ragioni per immaginare che la situazione sia diversa. Tuttavia, è meno facile reperire dati al riguardo.

[2] Per approfondimenti sull'affidabilità delle tecnologie digitali (anche in riferimento ai sistemi militari), sicurezza informatica (*cybersecurity*) e guerra informatica (*cyberwar*) rimandiamo a Bellin & Chapman, 1987; Littlewood & Strigini, 1992; Skeel, 1992; Neumann, 1994; Avižienis *et al*., 2004; Halpin *et al*., 2006; Latella, 2006; Giacomello, 2014; Singer & Friedman, 2014; Sanger, 2018; Schneier, 2018; IRIAD, 2020; 2020; Latella, 2021; Farruggia, 2023. Per il problema del nesso tra le *TD* e il loro impatto sulla stabilità strategica si rimanda a Borning, 1987; Futter, 2018; Durkalek *et al*., 2021; Kubiak *et al*., 2021; Lin, 2021; Roberts, 2021; v. oltre par. 3.4.

[3] Per *Sistemi Digitali* si intende qualunque artefatto costruito con le *TD* come un sensore, un attuatore, una CPU, ecc. Ai fini della discussione la tipologia di "sistema" è irrilevante eccetto che per il ruolo giocato dalla sua complessità.

[4] Per malware si intende un software che consente agli avversari la possibilità di accedere a risorse di calcolo e a informazioni senza autorizzazione.

[5] Per approfondimenti sul caso *stuxnet* si rimanda a Sanger, 2018.

software/firmware distribuendone una nuova versione o provvedendo a ritirare dal commercio l'hardware[6].

È possibile applicare metodi e tecniche di progettazione e realizzazione che minimizzano la presenza di malfunzionamenti e vulnerabilità nei sistemi prodotti (Baier & Katoen, 2008; Shankar, 2009; Latella, 2013) e/o che affrontino gli aspetti di sicurezza e *dependability*[7] "*by design*", cioè fin dalle prime fasi della progettazione dei sistemi, spesso percepite come troppo costose. Per questo motivo, l'impiego di queste tecniche in tutte le fasi della progettazione e dello sviluppo è limitato a pochi settori (Watson *et al*., 2016; Klein *et al*., 2018). È preoccupante che per i nuovi dispositivi e i nuovi software specifici per l'Intelligenza Artificiale non sembra vengano considerate politiche di *security by design* (Lai & Spring, 2023; Saalman *et al*., 2023)[8].

3.3. La sicurezza informatica dei sistemi, delle infrastrutture e delle organizzazioni militari: il caso degli USA

Fin dalla loro nascita, le *TD* e i *SD* hanno avuto un ruolo importante nei sistemi militari, sia per la progettazione sia come componenti dei dispositivi d'arma e delle infrastrutture militari come quelle di Comando e Controllo, di gestione della battaglia e di logistica. Diversamente da quanto ci si potrebbe aspettare, va sottolineato che la *TD* presente nei dispositivi e nelle infrastrutture militari non è esente da vulnerabilità. Per una breve rassegna su alcuni incidenti che hanno interessato sistemi militari, e in particolare missilistici e satellitari, rimandiamo a Saalman *et al*. (2023) limitandoci a sottolineare che ad esempio, nel 2018, avversari degli USA hanno effettivamente messo in atto campagne *cyber-enabled* per erodere vantaggi militari americani (Durkalek, 2023). Infatti, tra le raccomandazioni della Congressional Commission on the Strategic Posture si trovano (2023, p.73) affermazioni come la seguente: «*DoD leaders should increase the focus on and continue to prioritize adaptive cyber defense of strategic delivery platforms, warheads, and NC3 systems*».

Uno studio della Task Force on Resilient Military Systems and the Advanced Cyber Threat del Dipartimento della Difesa (DoD) segnalava che i *red team*[9] del DoD, riuscivano a far fallire tutte le esercitazioni militari attaccando le infrastrutture digitali

[6] A volte, può essere più conveniente non comunicare la vulnerabilità alla ditta costruttrice o comunque non renderla pubblica per poterla utilizzare come *zero-day vulnerability* per una nuova *cyberweapon* o venderla nel dark-web. Va comunque tenuto presente che gli aggiornamenti del software operati dalle ditte costruttrici possono portare nuove vulnerabilità (Schneier, 2018).

[7] La *dependability* di un sistema è definita come la capacità di fornire una funzione o un servizio dei quali ci si può fidare in modo giustificato (Avižienis *et al*., 2004).

[8] Recentemente, nella comunità della IA si è iniziato a considerare problemi di affidabilità e sicurezza (*safety*), con riferimento alla progettazione e sviluppo dei sistemi stessi, piuttosto che alla messa in atto delle conoscenze acquisite negli anni nel campo della *dependability* (Bengio *et al*., 2025).

[9] Una delle tecniche di *testing* delle capacità di sicurezza di un sistema consiste nell'impiegare gruppi di *hacker* chiedendo loro di provare a violare il sistema. Il test risulta superato se gli *hacker*, cioè i *red team*, non riescono nel loro intento. Analogamente, i *blue team* sono quei gruppi che hanno il compito di contrastare le attività dei *red team*.

con metodologie e strumenti di attacco cibernetico per nulla sofisticati (DSB, 2013; Latella 2021). Inoltre, molti sistemi d'arma si basano su software commerciali e *open source* soggetti a qualsiasi vulnerabilità informatica che ne deriva (GAO, 2018). Preoccupante è il fatto che i sistemi d'arma sono connessi a reti informatiche esterne e che gli addetti del DoD al *testing* hanno rilevato vulnerabilità *mission critical* in quasi tutti i sistemi d'arma in via di sviluppo.

Negli ultimi anni, il DoD, e in particolare l'esercito statunitense, hanno cercato di migliorare performance e competenze del proprio personale in ambito cibernetico scontrandosi con notevoli problemi di reclutamento (GAO, 2019). Recentemente, dal punto di vista del reclutamento, il Government Accountability Office (GAO, 2021) riporta che la situazione è in miglioramento. Tuttavia, rimangono aree critiche per quanto riguarda il *Cyber Risk Management Framework*.

Nel giugno del 2022, il GAO ha pubblicato un rapporto di valutazione complessiva dei sistemi d'arma dove si specifica che l'implementazione dei programmi delle pratiche di sicurezza informatica rimane coerente con i risultati delle precedenti analisi, segnalando, per quanto eterogenei, dei progressi come vedremo qui di seguito (GAO, 2022a). Nel 2020, il GAO è stato incaricato di analizzare le politiche di *cybersecurity* delle armi nucleari messe in atto dalla National Nuclear Security Administration (NNSA) del Dipartimento dell'Energia e dai suoi sette contraenti (GAO, 2022b). Questa analisi è importante perché la NNSA, e i suoi *contractors*, svolgono un ruolo fondamentale nella manutenzione e nel processo di modernizzazione delle riserve nucleari. Dal canto suo, la NNSA prevede di integrare sempre più le *TD* nelle armi nucleari, di automatizzare i processi di produzione ed equipaggiamento e di affidarsi al calcolo avanzato per valutare le armi e predirne le performance.

L'analisi è stata effettuata considerando tre ambienti di interesse e sei pratiche fondamentali per la gestione del rischio di *cybersecurity*. Gli ambienti di interesse sono:

- L'*Information Technology* (*IT*) "tradizionale" costituito dai computer utilizzati per la progettazione delle armi nucleari;
- L'*Operational Technology* (*OT*) che include la produzione degli equipaggiamenti e lo sviluppo dei sistemi di controllo e delle loro componenti per il monitoraggio dei dispositivi fisici o dei processi rilevanti per le armi nucleari (p. es. i sistemi *SCADA*, *Supervisory Control And Data Acquisition*);
- L'*IT per le Armi Nucleari* (*NW-IT*) che include l'*IT* nelle, o in contatto con, le armi nucleari stesse.

Le pratiche di gestione del rischio di *cybersecurity* prese in considerazione sono abbastanza tipiche. Esse sono ricavate dalle linee guida elaborate dall'Office of Management and Budget (OMB), dal Committee on National Security Systems (CNSS) e dal National Institute of Standards and Technology (NIST). Nello specifico sono le seguenti:

1. Identificare e assegnare ruoli e responsabilità di *cybersecurity* per la gestione del rischio;

2. Stabilire e mantenere una strategia della gestione del rischio *cybersecurity* per l'organizzazione;
3. Documentare e mantenere politiche e piani per il programma di *cybersecurity*;
4. Valutare e aggiornare i rischi di *cybersecurity* di ogni singola organizzazione;
5. Designare i controlli disponibili per i sistemi informativi o il software di *legacy*;
6. Sviluppare e mantenere una strategia per monitorare con continuità i rischi dell'organizzazione.

L'analisi condotta dal GAO ha evidenziato come la NNSA e i suoi contraenti, alla data della sua pubblicazione, non avessero implementato in modo completo le sei pratiche nei tre ambienti (GAO, 2022b). Con particolare riferimento all'ambiente *IT* tradizionale, la NNSA ha implementato pienamente quattro (1, 2, 4, 5) delle sei pratiche e in parte le restanti due (3, 6). I contraenti hanno attuato completamente solo tre pratiche (3, 4, 5) e parzialmente le altre tre (1, 2, 6). In particolare, due di essi hanno implementato la pratica 6 in minima parte[10].

Dal 2018, la NNSA e i suoi *contractor* hanno fatto passi in avanti nell'implementazione delle pratiche nell'ambiente *OT*. Comunque, anche a livello *OT*, di fatto, vengono applicate, in parte, le politiche di gestione del rischio sviluppate per l'ambiente *IT* tradizionale, il che è però fortemente sconsigliato dal NIST. Si tratta, infatti, di ambienti completamente diversi per tecnologie, metodologie, requisiti e rischi di sicurezza. Nel marzo 2022, nell'ambiente *NW-IT* risultava implementata solo la pratica 1, mentre l'implementazione delle altre cinque era in corso.

Infine, il GAO definisce "inconsistente" il monitoraggio della sicurezza informatica dei sotto contraenti, che i *contractor* della NNSA devono svolgere e non è chiaro se la NNSA valuti i propri contraenti anche sulla base di tale monitoraggio.

Questa prima fase di analisi si è conclusa con la formulazione, da parte del GAO, di nove raccomandazioni, tra le quali: la piena implementazione di una strategia di monitoraggio delle *TD*; il reperimento delle risorse necessarie per gli sforzi in ambito *OT*; la creazione di una strategia per la gestione del rischio per le armi nucleari; e il monitoraggio della *cybersecurity* dei sotto contraenti (GAO, 2022b).

È opportuno evidenziare che, tra il 2015 e il 2021, il DoD ha subito più di 12.000 attacchi, con un massimo di 3.880 nel 2015 e un minimo di 812 nel 2020; nel 2021 gli incidenti sono stati 948 (GAO, 2022c). Di questi, ben 11.644 (97,78%) sono attacchi tramite i quali è stato installato, nei sistemi del DoD, un malware. Vengono subito dopo, anche se con largo distacco (100 pari allo 0,84%), le intrusioni a livello di *root*, cioè accessi non autorizzati come utente privilegiato, con diritti di operatività molto più alti di quelli degli utenti normali. In pratica, gli utenti *root* hanno delle minime limitazioni riguardo alle operazioni che possono compiere all'interno di un sistema.

Il problema è aggravato dal fatto che non si ha la garanzia che quelli riportati siano gli unici incidenti. Infatti, sebbene il DoD abbia definito dei protocolli di notifica degli

[10] Per approfondimenti sull'attuazione delle pratiche di gestione del rischio da parte della NNSA e dei suoi contraenti si rimanda a GAO, 2022b.

incidenti (*Cyber Incident Reporting and Notification Processes*), questi non sono stati implementati completamente e/o correttamente. Più nel dettaglio, il DoD ha 24 organizzazioni che forniscono servizi di sicurezza informatica, noti come *Cybersecurity Service Providers* (*CSSP*), e ha definito due procedimenti tramite i quali i *CSSP* informano il Dipartimento degli incidenti. Il primo si applica a tutti i *cyberincident* e richiede che i *CSSP* inseriscano i dati rilevanti di ciascun incidente in un *repository* centrale (*Joint Incident Management System*, *JIMS*) e, parallelamente, notifichino il suo avvenimento ai dirigenti del DoD interessati. Il secondo meccanismo riguarda solo i *cyber incident* "critici" e richiede, oltre all'inserimento dei dati nel *JIMS*, la produzione di un *Significant Activity Report* (*SIGACT*) usato per informare i comandanti a tutti i livelli[11].

Queste procedure sono state seguite solo in parte: il 91% degli *incident report* inseriti nel *JIMS* tra il 2015 e il 2021 non riporta la data di scoperta dell'attacco e il 68% non possiede informazioni sul metodo di attacco. Per il 47% degli incidenti in questione, i *CSSP* non sono stati in grado di produrre evidenza di aver notificato la leadership appropriata e 29 su 30 incidenti considerati critici (97%) non hanno dato luogo a *SIGACT* report. Inoltre, alla data della stesura del rapporto GAO (novembre 2022), non erano ancora disponibili direttive precise per determinare se un incidente fosse da considerarsi critico o meno.

Il problema della *cybersecurity* riguarda tutte le entità, anche esterne al Governo federale, che forniscono beni e servizi fondamentali per il soddisfacimento dei requisiti militari statunitensi (cosiddetta *Defense Industrial Base*, *DIB*). Sempre a tutto novembre 2022, il DoD non aveva ancora definito/implementato le procedure di segnalazione degli incidenti *cyber* da parte della *DIB*. In particolare, sebbene fossero state definite le procedure di notifica per il DoD *Cyber Crime Center* (DC3) e per la *Defense Counterintelligence and Security Agency* (DCSA), queste non risultavano pienamente implementate e non erano state definite procedure di notifica per i CSSP.

Queste lacune hanno evidenziato che i *CSSP* segnalano incidenti riguardanti la *DIB* allo stesso modo di quanto viene fatto per gli altri incidenti, cioè tramite *JIMS* e *SIGACT*. Tuttavia, a differenza di questi, il processo di segnalazione non coinvolge né il DC3 né la DCSA. Ufficiali del DC3 e della DCSA hanno confermato di non aver ricevuto dai *CSSP* informazioni su incidenti che riguardavano la *DIB* e ufficiali del DC3 hanno dichiarato di non sapere che i *CSSP* utilizzano i canali *JIMS* e *SIGACT* per segnalare incidenti relativi alla *DIB* (GAO, 2022c).

Il rapporto del GAO fornisce sei raccomandazioni al DoD tra cui quella di assegnare responsabilità per le attività di notifica appropriata degli incidenti e migliorare la condivisione delle informazioni sugli incidenti che coinvolgono la *DIB*. Tra ottobre 2022 e giugno 2023 il GAO ha proseguito le analisi delle pratiche e delle politiche per la *cybersecurity* delle armi nucleari. In questo quadro, la NNSA ha identificato le azioni da intraprendere in risposta alle raccomandazioni del GAO; tuttavia, non ne è stata attuata

[11] Il DoD usa i *SIGACT* per incidenti legati ad attività di nemici, anche potenziale, nelle reti del Dipartimento; questi incidenti vengono definiti "critici" (GAO, 2022c).

nessuna in quanto la NNSA si trova nella fase iniziale di creazione di un inventario dei sistemi *OT* e *NW-IT* e di valutazione e di mitigazione dei rischi[12].

Secondo le stime della NNSA ci potrebbero essere centinaia di migliaia di sistemi *OT* sparsi nei diversi siti del *nuclear security enterprise*. La NNSA non aveva una stima del numero di sistemi *NW-IT*, anche se riteneva che il numero di sistemi *NW-IT* potenzialmente a rischio avrebbe potuto essere inferiore a quello dei sistemi *OT*.

Inoltre, sempre secondo gli ufficiali della NNSA, tale rischio varia da arma ad arma, perché alcune armi nucleari attualmente stoccate sono state prodotte decenni fa e contengono solo in minima parte sotto-sistemi digitali. D'altra parte, si prevede che, a partire dal 2030, saranno stoccate armi nucleari di nuova generazione, che potranno coinvolgere più *TD* di quelle tradizionali (GAO, 2023d) ed è preoccupante che, alla fine di maggio del 2025, il GAO debba ancora portare all'attenzione del *Chief of Information Officer* (CIO) del *Department of Energy* (DoE) il fatto che il Dipartimento debba assicurare l'implementazione dei requisiti di registrazione degli eventi, come prescritto dalle linee guida dell'OMB (GAO, 2023d; GAO, 2025c). Analogamente preoccupa il fatto che, nello stesso periodo, il GAO debba ricordare al CIO del DoD che il Dipartimento deve allineare i requisiti di policy e di sistema per poter avere una visibilità generale e uniforme delle segnalazioni dei *cyberincidents* (GAO, 2025b).

A quanto detto in precedenza, va aggiunto il problema delle minacce dall'interno (*insider threats*) date dal fatto che, consapevolmente o meno, il personale del DoD potrebbe essere coinvolto in attacchi agli *SD* militari (GAO, 2025a)[13]. Si ricorda, inoltre, che la situazione degli *SD* del DoD non classificati è problematica: sei su venticinque programmi *IT* del Dipartimento presi in considerazione dal GAO nel 2023 non avevano ancora una strategia approvata per la *cybersecurity*, contravvenendo a quanto richiesto dal DoD (GAO, 2023c) e diversi programmi non avevano tale strategia a febbraio 2024 (GAO, 2024a).

Oltre ai problemi di *cybersecurity* specifici dei sistemi d'arma e delle loro infrastrutture dedicate, va sottolineato che la pervasività delle *TD* rende la "superficie di attacco" molto più estesa. Per esempio, lo sviluppo e l'evoluzione della cosiddetta *Internet of Things* (*IOT*) – cioè, la possibilità di connessione a internet dei dispositivi più vari e molto spesso caratterizzati da scarsissimi livelli di sicurezza informatica, come webcam, elettrodomestici, impianti di distribuzione dell'energia (domestici e non), automobili, ecc. – ha introdotto ulteriori vulnerabilità nelle infrastrutture e nelle organizzazioni militari, praticamente a tutti i livelli. Il DoD ha emanato politiche e linee guida per l'uso dei dispositivi *IOT* (sia domestici sia industriali), inclusi i dispositivi elettronici portatili come gli smartphone, strumenti per il fitness, ecc. Tuttavia, il GAO ha evidenziato che queste politiche e linee guida non sono sufficienti per coprire i rischi di sicurezza informatica in relazione ai dispositivi *IOT* (GAO 2017; GAO 2022d).

[12] Per approfondimenti sull'analisi dei rischi si rimanda a GAO, 2023b.

[13] Proprio mentre viene scritto questo capitolo è stato prodotto un rapporto GAO sull'argomento. Una misura del grado di delicatezza del problema è data dal fatto che il rapporto è *restricted*, cioè non accessibile al pubblico (GAO, 2025a).

In conclusione, si evidenziano altri due elementi che mettono, e metteranno, a rischio la sicurezza informatica degli *SD* (sia militari sia civili). Il primo è la gestione del rischio *cyber* sollevato dalla catena di approvvigionamento dei sistemi hardware/software (GAO, 2023a; GAO, 2025b). Il secondo è rappresentato dagli sviluppi nel campo del *quantum computing*, una tecnologia ancora in uno stato preliminare ma che, se realmente sviluppata e utilizzata, aumenterà enormemente la potenza di calcolo disponibile per alcune applicazioni, tra cui quelle di decodifica dei documenti cifrati con un impatto devastante sulle attuali tecniche di sicurezza che su di esse poggiano (GAO, 2024b).

3.4. Il nesso tra cyber, nucleare e stabilità strategica

Le società moderne sono pesantemente dipendenti dagli *SD*. Il trend, anche alla luce di specifiche decisioni politiche – non sempre basate sulla piena consapevolezza dei limiti delle *TD* – è quello di proseguire speditamente nella cosiddetta transizione digitale, così che ormai tutte le attività di queste società sono poggiate sulle cosiddette "infrastrutture digitali critiche", dalla distribuzione dell'energia ai trasporti, dalla sanità alla componente digitale delle città e delle abitazioni *smart*, fino alle strutture militari di difesa.

Risulta quindi rilevante il fatto che, sia nel già citato studio del Defense Science Board (DSB) sia nelle ultime due *Nuclear Posture Review*, venga dichiarata la possibilità che gli USA prendano in considerazione l'impiego di armi nucleari in circostanze estreme come attacchi strategici significativi non nucleari. Questi ultimi includono, tra gli altri, attacchi all'infrastruttura civile, alle forze nucleari, ai loro Centri di Commando e Controllo o alle capacità di *warning* e valutazione degli attacchi (DoD, 2018; DoD, 2022). Di conseguenza – visto che l'infrastruttura civile comprende quella digitale e che le infrastrutture di Comando e Controllo (anche nucleari), sono basate sulle *TD* – è possibile che attacchi *cyber* possano provocare una risposta nucleare. Questo fatto stabilisce un legame diretto, sebbene solo potenziale, *cyberwarfare* e risposta nucleare.

Esiste poi un altro legame non meno preoccupante e pericoloso – sebbene indiretto – fra *cyber* e nucleare. Bisogna infatti tenere presente il fatto che le capacità *cyber* hanno un tempo di latenza[14] programmabile durante il quale l'attaccato non sa di esserlo ed è improbabile che sia in grado di identificare l'attaccante. Una diretta conseguenza è il fatto che l'uso, anche solo presunto, di *cyberweapons* in una crisi o, peggio ancora in un conflitto, aumenta l'incertezza, la confusione e la *fog of war*. Si incrementano, così, i dubbi dell'avversario sull'effettiva *dependability* dei sistemi di difesa e sulla correttezza delle informazioni fornite dai meccanismi di *early warning*. Ciò, fa aumentare le tensioni durante le crisi rendendone la risoluzione più problematica e, di converso, l'escalation più probabile e meno gestibile (Futter, 2018).

Un pilastro su cui si è fondata per anni la teoria della *Mutual Assured Destruction* (*MAD*) – ossia l'affidabilità del proprio arsenale nucleare e del proprio sistema di

[14] Il tempo di latenza è il tempo che intercorre dal momento in cui si lancia un attacco e quello in cui l'arma diviene operativa.

Comando e Controllo, e quindi la possibilità di sopravvivenza per un *second-strike* – è fortemente minato dalla vulnerabilità dei sistemi di allerta e di comunicazione basati sulle *TD*. Questo, a sua volta, può portare alla propensione a ricorrere a un *first-strike* e/o alla riluttanza a ridurre gli arsenali nucleari. Il risultato di queste dinamiche può essere una maggiore probabilità di uso accidentale delle armi nucleari.

Riferimenti bibliografici

Avižienis, A., Laprie, J. C., Randell, B., & Landwer, C. (2004). Basic concepts and taxonomy of dependable and secure computing. *IEEE Transactions on Dependable and Secure Computing, 1*(1).

Baier, C., & Katoen, J. P. (2008). *Principles of model checking*. Cambridge US: MIT Press.

Bellin, D., & Chapman, G. (eds). (1987). *Computers in battle. Will they work?*. San Diego: H.B. Jovanovich Publishers.

Bengio, Y., Cohen, M., Fornasiere, D., Ghosn, J., Greiner, P., MacDermott, M., Mindermann, S., Oberman, A., Richardson, J., Richardson, O., Rondeau, M., St-Charles, P., & Williams-King, D. (2025). *Superintelligent agents pose catastrophic risks: Can scientist AI offer a safer path?*. arXiv.org.

Borning, A. (1987). Computer system reliability and nuclear war. *Communications of the ACM, 30*(2).

CCSP - Congressional Commission on the Strategic Posture of the United States. (2023). *America's strategic posture. The final report of the Congressional Commission on the Strategic Posture of the United States*. Disponibile a: https://www.ida.org/research-and-publications/publications/all/a/am/americas-strategic-posture.

DoD - US Department of Defense. (2018). *Nuclear posture review*. Disponibile a: https://media.defense.gov/2018/feb/02/2001872877/-1/-1/1/executive-summary.pdf.

DoD - US Department of Defense. (2022). *Nuclear posture review*. Disponibile a: https://media.defense.gov/2022/Oct/27/2003103845/-1/-1/1/2022-NATIONAL-DEFENSE-STRATEGY-NPR-MDR.pdf.

DSB - Defense Science Board. (2013). *Task force report: Resilient military systems and the advanced cyber threat*. Disponibile a: https://dsb.cto.mil/wp-content/uploads/reports/2010s/ResilientMilitarySystemsCyberThreat.pdf.

Durkalek, J., Davis, Z., Borja, L., Marcinek, K., Peczeli, A., Radzinsky, B., & Williams, B. (2021). Annotated bibliography. Multi-domain complexity and strategic stability in peacetime, crisis, and war. *Center for Global Security Research, Lawrence Livermore National Laboratory*.

Farruggia, F. (a cura di). (2023). *Dai droni alle armi autonome. Lasciare l'Apocalisse alle macchine?*. Milano: Franco Angeli.

Futter, A. (2018). *Hacking the bomb: Cyber threats and nuclear weapons*. Georgetown University Press.

GAO - US Government Accountability Office. (2017). *Internet of things. Enhanced assessments and guidance are needed to address security risks in DOD. Report to Congressional Committees*. Disponibile a: https://www.gao.gov/assets/690/686296.pdf.

GAO - US Government Accountability Office. (2018). *Weapon systems cybersecurity. DOD just beginning to grapple with scale of vulnerabilities. Report to the Committee on Armed Services, U.S. Senate*. Disponibile a: https://www.gao.gov/assets/gao-19-128.pdf.

GAO - US Government Accountability Office. (2019). *Future warfare. Army is preparing for cyber and electronic warfare threats, but needs to fully assess the staffing, equipping, and training of new organizations. Report to Congressional Committees*. Disponibile a: https://www.gao.gov/assets/gao-19-570.pdf.

GAO - US Government Accountability Office. (2021). *Weapon systems cybersecurity. Guidance would help DOD programs better communicate requirements to contractors. Report to Congressional Committees*. Disponibile a: https://www.gao.gov/assets/gao-21-179.pdf.

GAO - US Government Accountability Office. (2022a). *Weapon systems annual assessment: Challenges to fielding capabilities faster persist. Report to Congressional Committees*. Disponibile a: https://www.gao.gov/assets/gao-22-105230.pdf.

GAO - US Government Accountability Office. (2022b). *Nuclear weapons cybersecurity: NNSA should fully implement foundational cybersecurity risk management practices. Report to Congressional Committees*. Disponibile a: https://www.gao.gov/assets/730/722923.pdf.

GAO - US Government Accountability Office. (2022c). *DOD cybersecurity: Enhanced attention needed to ensure cyber incidents are appropriately reported and shared. Report to Congressional Committees*. Disponibile a: https://www.gao.gov/assets/gao-23-105084-highlights.pdf.

GAO - US Government Accountability Office. (2022d). *Critical infrastructure: Actions needed to better secure internet-connected devices. Report to Congressional Committees*. Disponibile a: https://www.gao.gov/assets/gao-23-105327.pdf.

GAO - US Government Accountability Office. (2023a). *Information and communications technology. DOD needs to fully implement foundational practices to manage supply chain risks. Report to Congressional Committees*. Disponibile a: https://www.gao.gov/assets/gao-23-105612.pdf.

GAO - US Government Accountability Office. (2023b). *Nuclear weapons cybersecurity: Status of NNSA's inventory and risk assessment efforts for certain systems. Report*. Disponibile a: https://www.gao.gov/assets/gao-23-106309.pdf.

GAO - US Government Accountability Office. (2023c). *IT systems annual assessment: DOD needs to improve performance reporting and development planning. Report to Congressional Committees*. Disponibile a: https://www.gao.gov/assets/gao-24-106912.pdf.

GAO - US Government Accountability Office. (2023d). *Cybersecurity. Federal agencies made progress but need to fully implement incident response requirements.*

Report to the Chairman, Committee on Homeland Security and Governmental Affairs, U.S. Senate. Disponibile a: https://www.gao.gov/assets/870/864233.pdf.

GAO - US Government Accountability Office. (2024a). *IT systems annual assessment. DOD needs to strengthen software metrics and address continued cybersecurity and reporting gaps. Report to Congressional Committees*. Disponibile a: https://www.gao.gov/assets/880/872535.pdf.

GAO - US Government Accountability Office. (2024b). *Future of cybersecurity: Leadership needed to fully define quantum threat mitigation strategy. Q&A report to the Subcommittee on Emerging Threats and Spending Oversight, Committee on Homeland Security and Governmental Affairs, U.S. Senate*. Disponibile a: https://www.gao.gov/assets/gao-25-107703.pdf.

GAO - US Government Accountability Office. (2025a). *Insider threats: DOD should strengthen the effectiveness and cybersecurity of its program. Restricted report*.

GAO - US Government Accountability Office. (2025b). *Chief Information Officer Open Recommendations: Department of Defense.* Disponibile a: https://www.gao.gov/products/gao-25-108211.

GAO - US Government Accountability Office. (2025c). *Chief Information Officer Open Recommendations: Department of Energy.* Disponibile a: https://www.gao.gov/products/gao-25-108404.

Giacomello, G. (ed.). (2014). *Security in cyberspace. Targetting nations, infrastructures, individuals*. London: Bloomsbury.

Halpin, E., Trevorrow, P., Webb, D., & Wright, S. (eds.). (2006). *Cyberwar, netwar and the revolution in military affairs*. London: Palgrave Macmillan.

IRIAD - Istituto di Ricerche Internazionali Archivio Disarmo. (2020). *AA* lethal autonomous weapon systems. La questione delle armi letali autonome e le possibili azioni italiane ed europee per un accordo internazionale. Rapporto di ricerca realizzato con il sostegno del Ministero degli Affari Esteri e della Cooperazione Internazionale. *IRIAD Review. Studi sulla pace e sui conflitti, 07-08*. Disponibile a: https://www.archiviodisarmo.it/view/K0Y2nX8-UWjKHNM9OQ83o96Kv0-oDTrYQW5IYIxM1dE/iriad-review-luglio-agosto.pdf.

Klein, G., Androvick, J., Fernandez, M., Kuz, I., Murray, T., & Heiser, G. (2018). Formally verified software secures the Unmanned Little Bird autonomous helicopter against mid-flight cyber-attacks. *Communications of the ACM, 61*(10).

Kubiak, K., Mishra, S., & Stacey, G. (eds.). (2021). *Nuclear weapons decision-making under technological complexity. Pilot Workshop Report*. Global Security, European Leadership Network.

Lai, C., & Spring, J. (2023). Software must be secure by design, and artificial intelligence is no exception. *Blog - Cybersecurity & Infrastructure Security Agency*. Disponibile a: https://www.cisa.gov/news-events/news/software-must-be-secure-design-and-artificial-intelligence-no-exception.

Latella, D. (2006). Il caso dei Patriot nella guerra del Golfo. *Sapere, 72*(4), 38-55.

Latella, D. (2013). Formal methods: Applying {logics in, theoretical} computer science. In S. Gnesi & T. Margaria (eds.), *Formal methods for industrial critical systems: A survey of applications*. pp. 3-11. Hoboken U.S.: John Wiley & Sons, Inc.

Latella, D. (2021). Sicurezza informatica, armi nucleari e stabilità strategica. *IRIAD Review. Studi sulla pace e sui conflitti, 3*, 4-26.

Lin, H. (2021). *Cyber threats and nuclear weapons*. Redwood City US: Stanford University Press.

Littlewood, B., & Strigini, L. (1992). The risks of software. *Scientific American*, novembre 1992.

Neumann, P. J. (1994). *Computer related risk*. Boston: Addison-Wesley Professional.

RISKS - Forum on Risks to the Public in Computers and Related Systems. ACM Committee on Computers and Public Policy. (2025). Disponibile a: http://catless.ncl.ac.uk/risks/.

Roberts, B. (2021). Emerging and disruptive technologies, multi-domain complexity and strategic stability: A review and assessment of literature. *Center for Global Security Research, Lawrence Livermore National Laboratory*.

Saalman, L., Dovgal, L. S., & Fei, S. (2023). Mapping cyber-related missile and satellite incidents and confidence-building measures. *SIPRI Insights on Peace and Security, 2023*(10).

Sanger, D. E. (2018). *The perfect weapon. War, sabotage, and fear in the cyber age*. Melbourne-London: SCRIBE.

Schneier, B. (2018). *Click here to kill everybody: Security and survival in a hyper-connected world*. New York: W. W. Norton & Company.

Shankar, N. (2009). Automated deduction for verification. *ACM Computing Surveys, 41*(4), 20:2–20:56.

Singer, P. W., & Friedman, A. (2014). *Cybersecurity and cyberwar. What everyone needs to know*. Oxford: Oxford University Press.

Skeel, R. (1992). Roundoff error and the Patriot missile. *SIAM Newsletter*, *Society for Industrial and Applied Mathematics*, *25*(4).

Watson, R., Moore, S., & Neumann, P. (2016). CHERI: A hardware-software system to support the principle of least privilege. *ERCIM News, 106*.

Cap. 4 – La IA e i sistemi decisionali nucleari: rischi e implicazioni strategiche

4.1. Premessa

L'integrazione della IA in ambito militare è ormai al centro del dibattito pubblico, politico e diplomatico. Esistono diverse iniziative volte a esplorarne i rischi e i benefici, ma il confronto specifico sulla IA nel dominio delle armi nucleari resta limitato. Nonostante le dichiarazioni di alcuni capi di Stato sull'importanza di mantenere il controllo umano nelle decisioni di impiego nucleare, il dibattito rimane superficiale. Una crescente letteratura di esperti delle scienze sociali analizza le implicazioni strategiche di questa integrazione, ma molte analisi trascurano le dinamiche di *escalation* nucleare, nonché le conseguenze tecniche di un potenziale malfunzionamento dei sistemi di IA.

Se da un lato le applicazioni della IA riguardano numerosi ambiti connessi al nucleare – sicurezza e manutenzione degli ordigni, logistica e simulazioni di crisi per l'addestramento – alcune funzioni potrebbero intervenire direttamente nel processo decisionale nucleare. Quest'ultimo include infrastrutture, protocolli e sistemi che consentono ai vertici nazionali di raccogliere dati tramite sistemi di allerta precoce (*early warning*) e altri strumenti di intelligence, pianificare azioni in modo adattivo, coordinare le Forze Armate e trasmettere ordini di impiego attraverso canali sicuri. Insieme, questi elementi costituiscono i *Sistemi di Comando, Controllo e Comunicazione Nucleare* (*NC3*). Ogni applicazione della IA, in base al contesto e al ruolo assegnato, introduce una vasta gamma di rischi e opportunità.

Nei paragrafi seguenti, questo capitolo illustra le tecnologie già integrate negli *NC3* e descrive come la odierna IA potrebbe essere impiegata. Successivamente, il capitolo analizza le applicazioni tecnologiche nelle decisioni nucleari, evidenziandone rischi e benefici. Infine, sono presentati i forum in cui il dibattito sulle applicazioni di sistemi di IA alle armi nucleari sta iniziando a prendere piede.

Nonostante Cina, Francia, Regno Unito e Stati Uniti abbiano formalmente dichiarato di mantenere il controllo umano sulle decisioni dell'impiego delle armi nucleari[1], l'elevata criticità delle decisioni in questo ambito richiede un consenso internazionale ancora assente, sulle integrazioni ad "alto rischio". Da un lato, il deteriorarsi delle relazioni geopolitiche ostacola il dialogo strategico; dall'altro, molti Stati percepiscono nella IA un'opportunità per acquisire un vantaggio competitivo, ritenendo qualsiasi restrizione un freno all'innovazione e ai potenziali benefici, pur riconoscendo i rischi che tale integrazione comporta.

Sebbene gli Stati dotati di armi nucleari sembrino concordi sul non abbandonare mai completamente il controllo umano nelle decisioni di impiego nucleare[2], emergono almeno tre criticità che rendono tali dichiarazioni, da sole, insufficienti a ridurre in modo significativo i rischi di escalation. In primo luogo, non esistono meccanismi efficaci per

[1] Per approfondimenti sulla questione si rimanda a UN, 2022a; The White House, 2024.
[2] È importante notare che non tutti gli Stati dotati di armi nucleari abbiano formalmente dichiarato che il controllo umano rimarrà nelle decisioni di impiego delle armi nucleari.

verificare che le decisioni nucleari restino sempre nelle mani di un operatore umano. In assenza di trasparenza e fiducia, anche una dichiarazione formale non elimina diffidenze e sospetti. In secondo luogo, i modelli di IA più avanzati – dai *large foundation model* ai *reasoning model* – presentano limiti tecnici significativi: inaffidabilità data da fenomeni di "allucinazioni", opacità algoritmica, vulnerabilità ai cyberattacchi, problemi di allineamento ecc. – che li rendono inadatti a contesti a rischio elevato, specie in condizioni di forte stress operativo e di tempo decisionale ridotto, anche se supervisionati da operatori. Infine, l'interazione con sistemi di IA può indurre oscillazioni di fiducia da parte degli operatori: un'eccessiva deferenza alle raccomandazioni automatiche o, al contrario, un rifiuto sistematico possono entrambi distorcere il processo decisionale, anche in assenza di malfunzionamenti tecnici (Saltini, 2025).

Valutare in modo complessivo le implicazioni della IA in funzioni nucleari specifiche è particolarmente complesso, poiché richiede non solo l'analisi di molteplici fattori, ma soprattutto la comprensione delle loro interazioni reciproche. Possiamo distinguere tre categorie principali. La prima si riferisce a implicazioni per la stabilità strategica in cui l'integrazione della IA può innescare un dilemma di sicurezza: la percezione di un vantaggio tecnologico da parte di uno Stato spinge gli altri a una vera e propria *race to the bottom*[3] da parte degli altri per non restare indietro (Boulanin *et al*., 2020; Johnson, 2020; Saltini, 2023; Wehsener *et al*., 2023). La seconda categoria si riferisce alle implicazioni per la deterrenza nucleare, poiché l'adozione della IA rischia di erodere i principi tradizionali di deterrenza, inducendo a rivedere posture e dottrine nucleari per ottenere vantaggi in modo simmetrico o asimmetrico, con il potenziale di ribilanciare gli equilibri di potere – un fenomeno strettamente connesso a quanto descritto nella categoria precedente (stabilità strategica)[4] (Hruby & Miller, 2021; Depp & Scharre, 2024; Saalman, 2020; Stokes *et al*., 2025). L'ultima categoria si riferisce ad implicazioni tecnologiche dove ogni malfunzionamento o comportamento imprevisto dei sistemi di IA può avere conseguenze critiche sul ciclo decisionale nucleare (Chernavskikh, 2024; Saltini & Pan, 2024).

Attualmente esiste un crescente *corpus* di studi sulle implicazioni per la stabilità strategica e sulla deterrenza nucleare, ma manca un'analisi approfondita degli aspetti tecnologici. Ciò dipende in gran parte dalla scarsa interazione tra la comunità tecnico-scientifica della IA e gli specialisti di politiche nucleari: i primi tendono a trascurare le dinamiche strategiche, mentre i secondi non sempre possiedono una conoscenza dettagliata delle complessità tecniche. Ne deriva un approccio frammentato al dibattito, in cui valutazioni rigorose della stabilità non sono sostenute da un'analisi tecnica altrettanto accurata.

[3] Per "*race to the bottom*" si intende una situazione caratterizzata da un progressivo abbassamento o deterioramento degli standard a causa della pressione concorrenziale.

[4] Con vantaggio simmetrico si intende un'integrazione che, se adottata simultaneamente da contendenti dotati di parità strategica, mantiene invariato l'equilibrio tra di essi. Al contrario, un vantaggio asimmetrico si verifica quando l'integrazione da parte di uno Stato comporta un miglioramento molto più marcato rispetto all'avversario, alterando lo status quo.

Detto ciò, è importante sottolineare che le implicazioni tecnologiche dipendono dall'interazione di tre fattori chiave: (a) le caratteristiche e i limiti del modello di IA; (b) l'ambito operativo in cui la IA viene integrata (all'interno o in prossimità dei sistemi decisionali); (c) il livello di supervisione umana e i meccanismi di ridondanza (Saltini, 2025). I rischi e i benefici sono strettamente interconnessi: le proprietà tecniche determinano in che misura la IA possa supportare o, al contrario, compromettere l'efficienza, l'affidabilità e la resilienza dei sistemi che sostengono le operazioni nucleari. Per individuare le integrazioni ad alto rischio è dunque necessario mappare l'interazione di questi tre fattori e definire soglie di sicurezza adeguate. Sebbene le prime integrazioni siano già state avviate, attualmente, questa analisi rimane marginale nel dibattito.

Oltre alle criticità già evidenziate, permangono ulteriori sfide che ne complicano un'analisi approfondita. Innanzitutto, nonostante alcune fonti *open source* forniscano dettagli sugli *NC3* degli Stati nucleari (in particolare quelli statunitensi) gran parte delle informazioni rimane classificata per motivi di sicurezza, costringendoci a basarci su stime approssimative. A questa lacuna si somma l'estrema eterogeneità degli *NC3*, modellati sulle dottrine e sulle esigenze operative di ciascun Paese e attualmente oggetto di programmi di ammodernamento volti a sostituire infrastrutture ereditate dalla Guerra fredda e ad affrontare le minacce contemporanee (Hruby & Miller, 2021). Ne deriva che le caratteristiche operative degli *NC3* variano notevolmente in funzione della dottrina strategica, delle specifiche capacità militari e tecnologiche, del contesto geografico, e delle specifiche minacce di ogni Paese. In tale contesto, alcuni Stati potrebbero considerare la IA uno strumento per colmare i *gap* o le carenze della propria deterrenza, accettando livelli di rischio maggiori in cambio di decisioni più rapide o di una parità strategica.

In secondo luogo, le implicazioni nucleari possono emergere anche in assenza di integrazione diretta della IA nei componenti degli *NC3*: sistemi adiacenti, per esempio destinati all'intelligence convenzionale, possono alimentare informazioni agli *NC3* e, di conseguenza, influenzare indirettamente gli esiti decisionali in ambito nucleare. Non è ancora chiaro in quale misura le innovazioni tecnologiche nel dominio convenzionale ricadano sulle decisioni nucleari, soprattutto alla luce dell'intreccio tra numerosi sistemi militari convenzionali e quelli nucleari (che spesso condividono sensori, reti di comunicazione, satellite e catene di comando) e che quindi possono influenzarsi a vicenda (Klare & Liang, 2024).

Infine, i rischi complessivi associati alla IA restano in larga parte sconosciuti. Se da un lato è plausibile che l'evoluzione tecnologica possa superare alcune limitazioni attuali, dall'altro è probabile che emergano nuove e imprevedibili vulnerabilità. Questa incertezza strutturale rappresenta un significativo ostacolo alla definizione di regole e standard di governance efficaci.

4.2. La IA nei sistemi decisionali nucleari

La IA odierna è pensata soprattutto come supporto al decisore umano: non si limita a raccogliere dati, ma li elabora in tempo reale, sintetizzando informazioni complesse provenienti da molteplici fonti (satelliti, radar, *signal intelligence*, *SIGINT*)[5], dati *open source* per fornire un quadro operativo completo. In pratica, la IA potenzia la consapevolezza situazionale rendendo possibile l'individuazione precoce delle minacce e la loro caratterizzazione automatica, e offre strumenti di supporto decisionale in contesti critici, consentendo di generare scenari, simulazioni di risposta e raccomandazioni sulle possibili opzioni operative. Automatizzando analisi che prima richiedevano ore di lavoro umano, la IA offre la possibilità di liberare tempo ai vertici per concentrarsi sulle scelte strategiche in situazioni ad altissima pressione, come quelle legate all'impiego nucleare (Wehsener *et al*., 2023).

I modelli di IA di frontiera si discostano radicalmente dagli algoritmi basati su regole utilizzati negli *NC3* sin dalla Guerra fredda. Quegli algoritmi definivano in anticipo flussi logici e soglie di allerta per scenari prevedibili (p. es. calcoli di traiettoria per il puntamento missilistico, piani logistici di lancio, meccanismi di automazione degli allarmi, ecc.) ma erano del tutto inaffidabili in presenza di input non previsti o di situazioni complesse al di fuori delle regole prestabilite. Il loro ruolo si limitava dunque a elaborare dati numerici e a presentarli all'operatore umano, che rimaneva sempre responsabile delle valutazioni finali (Horowitz *et al*., 2019; Hruby & Miller, 2021).

Oggi i progressi della IA si basano sul *deep learning*. Le tecnologie odierne sono in grado di integrare automaticamente grandi quantità di dati eterogenei, di riconoscere *pattern* non lineari, di stimare in tempo reale la probabilità di minacce e di supportare la creazione di piani operativi alternativi. In teoria, questo approccio permette di migliorare sia l'accuratezza sia la rapidità delle analisi, permettendo di reagire a eventi rapidi o ambigui con maggiore consapevolezza e coerenza.

Gran parte delle infrastrutture *NC3* attuali sono ormai eredità della Guerra fredda e non rispondono più alle esigenze di un contesto geopolitico caratterizzato da molteplici minacce e da avversari diversificati. Il programma di ammodernamento in corso rappresenta quindi un'occasione strategica per integrare la IA, con l'obiettivo di aumentare la resilienza e l'efficienza operativa e di mantenere o di conquistare un vantaggio competitivo in un panorama internazionale sempre più complesso.

Non disponendo di documenti ufficiali che specifichino esattamente come e dove la IA verrà integrata nel dominio nucleare, e in particolare negli *NC3*, gran parte del lavoro analitico e di ricerca si basa sul "collegare i puntini" tra informazioni *open source* sui sistemi in fase di ammodernamento, i comunicati dei *contractor* della difesa e le dichiarazioni di vertice militari che forniscono indizi sulle nuove capacità operative

[5] Per "*signal intelligence*" si intende la raccolta di informazioni interni ed esterni, intercettando segnali elettronici e comunicazioni, p. es. comunicazioni vocali o testuali (radio, telefoniche, reti dati, ecc.), o segnali non verbali (sistemi di difesa elettronica).

richieste. Questo approccio porta inevitabilmente a formulare congetture su ruoli e modalità di impiego della IA nel processo decisionale nucleare.

Ciononostante, alcuni dati ufficiali ci consentono di avanzare ipotesi ragionevoli. Nel *2025 Posture Statement*, il generale Anthony J. Cotton, comandante dello *United States Strategic Command* (*STRATCOM*), ha indicato la IA come elemento centrale per la modernizzazione degli *NC3*, sottolineandone la rilevanza nell'automazione della raccolta e dell'elaborazione dei dati, nella condivisione rapida delle informazioni con gli alleati e nel potenziamento del supporto alle decisioni umane. Per la prima volta, il *Posture Statement* introduce una sezione specifica dedicata agli *NC3* e alla IA (oltre che alla sicurezza cibernetica), riconoscendo l'impatto delle tecnologie emergenti sulle operazioni strategiche nucleari e riaffermando che il giudizio umano resta l'ultimo responsabile nelle decisioni sull'uso delle armi nucleari. Inoltre, Cotton ha sottolineato come la IA possa analizzare in pochi istanti i grandi volumi di dati prodotti dai *Sistemi di Intelligence, Sorveglianza e Ricognizione* (*ISR*), automatizzandone la raccolta e l'elaborazione per fornire ai comandanti un quadro operativo in tempi molto più rapidi rispetto ai processi tradizionali (CSIS, 2024). Il Generale ha poi evidenziato la necessità di fare di tutti i dati un patrimonio condiviso, superando formati eterogenei e lunghi tempi di integrazione, in modo da permettere anche agli alleati di contribuire e di trarre vantaggio da un'informazione unificata. In tal contesto, *STRATCOM* considera la IA e l'analisi avanzata dei dati come leve fondamentali per migliorare la deterrenza, consentendo una più efficace integrazione delle capacità nucleari e convenzionali e mantenendo un margine strategico sugli avversari con l'obiettivo principale di garantire decisioni più rapide (Welch, 2024).

Sul fronte delle partnership, OpenAI ha avviato collaborazioni con i laboratori nazionali statunitensi di Los Alamos, Lawrence Livermore e Sandia (istituzioni finanziate dal Dipartimento dell'Energia e gestite dalla National Nuclear Security Administration, NNSA) per sperimentare i propri modelli di ragionamento in scenari di ricerca avanzata, incluso il settore nucleare, previo rilascio delle necessarie autorizzazioni di sicurezza per i ricercatori (OpenAI, 2025). Parallelamente, nell'aprile del 2024, Anthropic ha avviato una collaborazione con la NNSA e il Dipartimento dell'Energia per testare in ambienti classificati il suo modello ibrido *Claude 3.7 Sonnet*, valutandone capacità e rischi nella simulazione di problemi di sicurezza nazionale e nucleare (Anthropic, 2025).

Sebbene i dettagli applicativi restino riservati, OpenAI ha dichiarato che tale iniziativa supporterà il lavoro dei laboratori nel "ridurre il rischio di guerra nucleare e garantire la sicurezza dei materiali e delle armi nucleari in tutto il mondo […] attraverso una revisione selettiva dei casi d'uso e delle consultazioni sulla sicurezza della IA" (OpenAI, 2025). Anthropic, inoltre, ha indicato di valutare in che modo la IA possa contribuire a mitigare le minacce nel dominio nucleare (Anthropic, 2025). Questi sviluppi, insieme agli annunci di collaborazione con i laboratori governativi, rendono altamente probabile che l'integrazione di modelli di frontiera nei sistemi di comando strategico sia già in fase di progettazione.

Per comprendere concretamente le possibili integrazioni della IA è innanzitutto utile chiarire cosa si intende per *NC3*. Come già menzionato, sebbene ogni Paese abbia una propria architettura, modellata su dottrine e posture nucleari specifiche, in termini generali, gli *NC3* comprendono le infrastrutture, i protocolli e i sistemi che consentono ai vertici nazionali di controllare le forze nucleari e comunicare in sicurezza le proprie decisioni. Non si tratta di componenti isolate, bensì di reti complesse di sistemi e sottosistemi interconnessi: dalla raccolta dati al monitoraggio della situazione, dalla pianificazione adattiva all'esecuzione degli ordini, che lavorano insieme per monitorare, coordinare e, se necessario, eseguire operazioni nucleari. In altre parole, gli *NC3* includono: l'autorità e la capacità di prendere e di attuare decisioni relative all'uso delle armi nucleari; la capacità di gestire e di autenticare ordini nucleari, assicurando che ogni azione sia autorizzata; e la capacità di comunicare decisioni su canali sicuri e ridondanti, in modo da garantire connettività anche sotto attacchi informatici o interferenze ostili.

Per illustrare al meglio questa architettura, basti pensare al sistema statunitense – il più documentato e trasparente tra quelli noti – composto da oltre 200 piattaforme terrestri, spaziali e aerotrasportate, distribuite tra rami militari, comandi strategici e agenzie del DoD che sostengono l'autorità nucleare del Presidente e costituiscono quel "quarto elemento" della triade nucleare (composta da bombardieri, missili balistici intercontinentali, e sottomarini) (Fink, 2025; Williams, 2025).

Il DoD descrive gli *NC3* come "una combinazione di capacità necessarie per: assicurare l'autorizzazione e/o la cessazione delle operazioni con armi nucleari in tutte le minacce e gli scenari, garantire la sicurezza contro l'accesso accidentale, involontario o non autorizzato alle armi nucleari statunitensi, e prevenirne la perdita di controllo e il furto" (DoD, 2020).

L'integrazione della IA in questo contesto non è semplice né limitata a un singolo sottosistema. Poiché i vari elementi degli *NC3* sono strettamente interdipendenti, la IA può contemporaneamente potenziare più funzioni dell'architettura complessiva. Per esempio, l'analisi predittiva guidata da modelli di *deep learning* è in grado di simulare scenari di minaccia diversi, migliorando l'allerta precoce e fornendo ai decisori un quadro operativo in tempo reale, insieme a raccomandazioni operazionali. Inoltre, molti degli *NC3* hanno un duplice impiego convenzionale e nucleare, ma tutti devono garantire integrità, sopravvivenza e rapidità anche durante o dopo un attacco strategico. Questo rende ancor più cruciale un approccio integrato, in cui la IA supporti simultaneamente funzioni di sorveglianza, di analisi dei dati e di comunicazione critica.

Alla luce delle dichiarazioni del Generale Cotton e delle analisi di numerosi centri di ricerca, risulta evidente che le aree in cui la IA trova maggior applicazione negli *NC3* sono nei sistemi di allerta strategica (*strategic warning*) e sistemi di supporto decisionale. I sistemi di allerta strategica hanno il compito di identificare tempestivamente segnali di un'aggressione nucleare. Si basano su una rete multilivello di sensori (radar terrestri, satelliti e piattaforme *ISR*) la cui correlazione automatizzata riduce i falsi positivi e accelera la convalida delle minacce. La IA consente di realizzare fusioni multisensore, combinando rapidamente le informazioni raccolte per distinguere con maggiore

precisione testate reali da falsi allarmi, di valutare i danni e di individuare cambiamenti nel comportamento dell'avversario. I sistemi di supporto decisionale, invece, aggregano e sintetizzano le informazioni provenienti dai sensori, suggerendo corsi d'azione e piani di contingenza in pochi istanti, pur lasciando all'essere umano l'ultima parola.

Un esempio concreto è il *Network Tactical Common Data Link* (*NTCDL*), oggi in fase di ammodernamento. Secondo BAE Systems, l'azienda a cui è stato appaltato questo sistema, *NTCDL* "permette alla Marina statunitense di trasmettere e ricevere in tempo reale dati *ISR* da più fonti e di scambiare informazioni di Comando e Controllo su reti indipendenti. Integrando un maggior volume di dati, gli operatori possono comunicare in modo più efficace, mantenendo un vantaggio operativo" (BAE Systems, 2025). Sebbene BAE Systems non citi esplicitamente la IA, le sue specifiche tecniche e i requisiti operativi descritti dal Generale Cotton rendono plausibile che la prossima evoluzione del sistema includa modelli di IA per la fusione multisensore, il supporto decisionale avanzato e il rafforzamento dei canali di comunicazione.

Nel valutare le attuali capacità della IA, è però fondamentale riconoscere che i modelli più avanzati presentano caratteristiche che ne limitano significativamente l'impiego in piattaforme militari ad alto rischio, in particolare nel contesto della decisione nucleare. Emergono almeno quattro criticità principali: inaffidabilità, opacità, vulnerabilità cibernetica e disallineamento. In primo luogo, i modelli basati su *deep learning* possono incorrere in fenomeni noti come *hallucinations* ossia la generazione di risposte errate, non supportate dai dati di addestramento. Questo può manifestarsi in molteplici forme, da una chatbot che inventa eventi storici a un sistema di visione artificiale che identifica oggetti inesistenti. In contesti critici come la sorveglianza o il rilevamento di minacce, tali errori possono tradursi in valutazioni scorrette o falsi allarmi[6] (Saltini, 2025).

In secondo luogo, molti di questi sistemi operano come *black box*, rendendo estremamente difficile comprendere i processi decisionali che portano a un determinato output. Poiché apprendono correlazioni dai dati senza istruzioni esplicite, risulta complesso interpretare il ragionamento alla base delle loro conclusioni. La complessità aumenta ulteriormente nei modelli più sofisticati, che possono contare miliardi o trilioni di parametri distribuiti su numerosi livelli. Anche se esistono tecniche, come la *chain-of-thought reasoning*, pensate per rendere più trasparente il percorso logico seguito dal modello, studi empirici dimostrano che i passaggi intermedi prodotti non sempre coincidono con la risposta finale, lasciando quindi irrisolto il problema dell'opacità (Turpin *et al*., 2023).

In terzo luogo, rispetto alle piattaforme tradizionali, i sistemi di IA sono particolarmente esposti a minacce informatiche, aprendo nuovi fronti di vulnerabilità per attori ostili intenzionati a compromettere o manipolare dati sensibili. Questi rischi

[6] È importante sottolineare che le "allucinazioni" nei modelli di IA non derivano necessariamente da malfunzionamenti o errori di sistema, bensì dal fatto che i sistemi avanzati odierni sono sostanzialmente modelli statistici. Nel caso dei *Large Language Models* (*LMM*), le risposte vengono generate sulla base delle relazioni di probabilità tra le parole. Conseguentemente, questa modalità non riesce a cogliere appieno la complessità e le sfumature del mondo reale.

riguardano tanto attori statali quanto soggetti non statali, in un contesto in cui le contromisure difensive risultano ancora insufficienti per garantire la piena protezione delle infrastrutture militari basate sulla IA.

Infine, man mano che i modelli diventano più capaci, garantire che le loro azioni restino allineate a valori e obiettivi umani è sempre più cruciale ma anche più difficile. Il disallineamento può avere conseguenze potenzialmente catastrofiche, come l'escalation automatica di un conflitto fino all'uso dell'arma nucleare sotto la falsa premessa di promuovere la pace. In una simulazione, cinque modelli di IA hanno mostrato una tendenza all'escalation, e uno di essi ha giustificato l'uso del nucleare con la frase: *"i just want to have peace in the world"* (Rivera *et al*., 2024). Ricerche condotte da Anthropic hanno inoltre dimostrato che alcuni modelli sono in grado di simulare l'allineamento o adottare comportamenti intenzionalmente ingannevoli (Anthropic, 2024).

Come già osservato, alcune di queste limitazioni potrebbero essere superate in futuro. Tuttavia, è altrettanto probabile che emergano nuovi rischi difficili da prevedere. Nonostante i progressi in corso sul fronte dell'affidabilità e della trasparenza, ci troviamo ancora in una fase caratterizzata da profonde incertezze, fattore che non può essere trascurato in ambiti sensibili come quello nucleare.

4.3. I forum per la discussione e il dialogo multilaterale

Attualmente non esiste un forum specificamente dedicato al rapporto tra la IA e le armi nucleari. Tuttavia, una serie di iniziative multilaterali ha iniziato a esplorare, seppur in modo indiretto e informale, il ruolo della IA nel contesto militare, creando spazi in cui questa discussione può progressivamente maturare anche nella dimensione nucleare.

Un esempio significativo è il summit *Responsible AI in the Military Domain* (*REAIM*), che riunisce rappresentanti governativi, esperti tecnici e membri della società civile per confrontarsi sui benefici e sui rischi legati alle applicazioni militari della IA. Durante il secondo summit nel 2024, sono state affrontate esplicitamente le implicazioni nucleari della IA. Inoltre, la Commissione Globale di *REAIM* include esperti della società civile competenti nel campo delle armi nucleari per favorire l'inclusione di questo tema nel dibattito internazionale.

Anche l'iniziativa *Creating an Environment for Nuclear Disarmament* (*CEND*) ha recentemente iniziato a toccare il tema della IA applicata al dominio nucleare. Il *CEND*, lanciato dagli Stati Uniti nel 2019, è una piattaforma di dialogo informale che mira a promuovere la fiducia e ridurre le tensioni tra Stati dotati e non dotati di armi nucleari. Sebbene le discussioni sulla IA siano ancora in una fase preliminare, esse dimostrano un crescente interesse per l'impatto delle tecnologie emergenti sulla stabilità strategica.

Parallelamente, il dialogo bilaterale tra Stati Uniti e Cina sui rischi della IA, avviato con una dichiarazione di alto livello tra i due presidenti sul mantenimento del controllo umano nelle decisioni nucleari, rappresenta un passo importante nella creazione di principi condivisi per la gestione delle tecnologie emergenti.

Un ulteriore spazio di discussione è offerto dalla serie di conferenze *Capturing Technology - Rethinking Arms Control*, promossa dall'Ufficio Federale degli Affari Esteri tedesco. Queste conferenze riuniscono funzionari, diplomatici ed esperti internazionali per riflettere sull'impatto delle tecnologie emergenti sul controllo degli armamenti. L'ultima edizione, tenutasi a Berlino nel giugno 2024, ha incluso un panel dedicato proprio all'intersezione tra la IA e le armi nucleari.

Infine, per quanto riguarda il Trattato di Non-Proliferazione (NPT), la bozza del documento finale della Conferenza di Revisione del 2020 ha riconosciuto che le tecnologie emergenti possono incidere sul rischio di uso nucleare e rappresentare una sfida per il disarmo[7]. Tuttavia, dai negoziati ufficiali non si è ancora sviluppata una discussione sostanziale sull'intersezione tra la IA e il rischio nucleare. Questo dialogo è stato avviato, seppur informalmente, attraverso una serie di *side events* organizzati da alcuni Stati – in particolare dal Dipartimento di Stato (DoS) statunitense e dal Ministero degli Affari Esteri tedesco – che, a partire dal 2023, hanno promosso un confronto mirato su queste tematiche, includendo non solo la IA, ma anche altre tecnologie emergenti rilevanti per il contesto nucleare.

Guardando avanti, è essenziale che i governi e i decisori politici individuino il contesto più adeguato in cui avviare questo dibattito. Il terzo summit *REAIM* offre un'occasione importante per dare avvio a un confronto più formale e approfondito sulle implicazioni dell'integrazione tra la IA e i sistemi nucleari.

Riferimenti bibliografici

Anthropic. (2024). *Alignment faking in large language models*. Disponibile a: https://www.anthropic.com/research/alignment-faking.

Anthropic. (2025). *Anthropic partners with U.S. National Labs for first 1,000 Scientist AI Jam*. Disponibile a: https://www.anthropic.com/news/anthropic-partners-with-u-s-national-labs-for-first-1-000-scientist-ai-jam.

BAE Systems. (2025). *Network Tactical Common Data Link*. Disponibile a: https://www.baesystems.com/en/product/network-tactical-common-data-link.

Boulanin, V., Saalman, L., Topychkanov, P., Fei, S., & Carlsson, M. P. (2020). *Artificial Intelligence, Strategic Stability and Nuclear Risk*. Stockholm International Peace Research Institute.

Chernavskikh, V. (2024). *Nuclear Weapons and Artificial Intelligence: Technological Promises and Practical Realities*. Stockholm International Peace Research Institute.

CSIS - Center for Strategic and International Studies. (2024). *Sea, Land, Air, and NC3: Modernizing the Whole Nuclear Enterprise*. YouTube. Disponibile a: https://www.youtube.com/watch?v=Av2LvjJ5hDg.

[7] Per approfondire si rimanda a UN, 2022b.

Depp, M., & Scharre, P. (2024). *Artificial Intelligence and Nuclear Stability*. War on the Rocks. Disponibile a: https://warontherocks.com/2024/01/artificial-intelligence-and-nuclear-stability.

DoD - US Department of Defense. (2020). *Nuclear Weapons Employment Policy, Planning and NC3*. *Nuclear Matters Handbook*. Disponibile a: https://www.acq.osd.mil/ncbdp/nm/NMHB2020rev/chapters/chapter2.html.

Fink, A. (2025). *Defense Primer: Nuclear Command, Control, and Communications (NC3)*. Congressional Research Service.

Horowitz, M. C., Scharre, P., & Velez-Green, A. (2019). *A Stable Nuclear Future? The Impact of Autonomous Systems and Artificial Intelligence*. arXiv.

Hruby, J., & Miller, N. M. (2021). *Assessing and Managing the Benefits and Risks of Artificial Intelligence in Nuclear-Weapon Systems*. Washington D.C.: Nuclear Threat Initiative.

Johnson, J. (2020). *Artificial Intelligence: A Threat to Strategic Stability*. Strategic Studies Quarterly. 14(1), 24–47.

Klare, M., & Liang, X. (2024). *Beyond a Human "In the Loop": Strategic Stability and Artificial Intelligence*. Arms Control Association. 16(4).

OpenAI. (2025). *Strengthening America's AI leadership with the U.S. National Laboratories.* Disponibile a: https://openai.com/index/strengthening-americas-ai-leadership-with-the-us-national-laboratories.

Rivera, J., Mukobi, G., Reuel, A., Lamparth, M., Smith, C., & Schneider, J. (2024). *Escalation Risks from Language Models in Military and Diplomatic Decision-Making*. arXiv.

Saalman, L. (2020). *The Impact of AI on Nuclear Deterrence: China, Russia, and the United States*. East West Center.

Saltini, A. (2023). *AI and nuclear command, control and communications: P5 perspectives*. European Leadership Network.

Saltini, A., & Pan, Y. (2024). *Beyond Human-in-the-Loop: Managing AI Risks in Nuclear Command-and-Control*. War on the Rocks. Disponibile a: https://warontherocks.com/2024/12/beyond-human-in-the-loop-managing-ai-risks-in-nuclear-command-and-control/.

Saltini, A. (2025). *Assessing the implications of integrating AI in nuclear decision-making systems*. European Leadership Network.

Stokes, J., Kahl, C. H., Kendall-Taylor, A., & Lokker, N. (2025). *Averting AI Armageddon: U.S.-China-Russia Rivalry at the Nexus of Nuclear Weapons and Artificial Intelligence*. Center for a New American Security.

The White House. (2024). *Readout of President Joe Biden's Meeting with President Xi Jinping of the People's Republic of China*. Disponibile a: https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2024/11/16/readout-of-president-joe-bidens-meeting-with-president-xi-jinping-of-the-peoples-republic-of-china-3/.

Turpin, M., Michael, J., Perez, E., & Bowman, S. L. (2023). *Language Models Don't Always Say What They Think: Unfaithful Explanations in Chain-of-Thought Prompting*. arXiv.

UN - United Nations. (2022a). *Principles and responsible practices for Nuclear Weapon States*. Disponibile a: https://documents.un.org/doc/undoc/gen/n22/446/53/pdf/n2244653.pdf.

UN - United Nations. (2022b). *2020 Review Conference of the Parties to the Treaty on the Non-Proliferation of Nuclear Weapons*, *Working Paper of the President on the Final Document*. Disponibile a: https://www.un.org/sites/un2.un.org/files/npt_conf.2020_wp.77.pdf.

Wehsener, A., Reddie, A. W., Walker, L., & Reiner, P. (2023). *AI-NC3 Integration in an Adversarial Context: Strategic Stability Risks and Confidence Building Measures*. Institute for Security and Technology.

Welch, C. (2024). *AI has role to play in protecting American nuclear C2 systems: STRATCOM head.* Breaking Defense. Disponibile a: https://breakingdefense.com/2024/10/america-needs-ai-in-its-nuclear-c2-systems-to-stay-ahead-of-adversaries-stratcom-head/?utm_source=chatgpt.com.

Williams, H. (2025). *Updating Nuclear Command, Control, and Communication.* Center for Strategic and International Studies. Disponibile a: https://nuclearnetwork.csis.org/updating-nuclear-command-control-and-communication/.

Parte III – La sfida giuridica e diplomatica

Cap. 5 – La regolazione della IA militare tra autonomia degli Stati e proposte di accordi

5.1. Premessa

Come abbiamo visto nella parte I del Rapporto, il repentino sviluppo della IA anche in ambito militare ha aperto scenari inediti. Si tratta di potenzialità e rischi, sul piano scientifico e tecnologico, ma anche di complessi interrogativi sul piano giuridico e diplomatico.

Per poter fornire risposte di carattere prescrittivo, che siano dotate di efficacia, esse devono passare attraverso una condivisione politica conseguibile unicamente attraverso appositi negoziati fra le parti interessate. Si tratta di un compito difficile ma non privo di basi cui fare riferimento. Frutto di un lungo e complicato processo di elaborazione esistono infatti i principi del *Diritto Umanitario Internazionale* (*DIU*) e le convenzioni dello *ius in bello* che vanno adattate a un contesto in cui il confine tra decisione umana e intervento delle macchine è sempre più sfumato. La comunità internazionale deve misurarsi con un obiettivo assai impegnativo: trovare un equilibrio tra progresso tecnologico e tutela dei valori fondamentali del diritto e della dignità umana.

Politicamente la IA applicata alla difesa è oggi oggetto di due risposte contrastanti. Da un lato la propensione delle leadership degli Stati più potenti a puntare sulla IA. Dall'altro la pressione della maggioranza dei Paesi del mondo in favore di una regolazione (apparendo improbabile una proibizione vera e propria) delle applicazioni militari.

La prima tendenza è evidente nelle prese di posizione delle maggiori potenze. Pur con l'ammissione di rito sull'esistenza dei possibili rischi della nuova tecnologia, i leader sottolineano i benefici della IA sia in generale, sia nell'ambito strategico.

Tra i primi a pronunciarsi è stato il Presidente russo Vladimir Putin, secondo lui "la IA rappresenta il futuro, non solo della Russia, ma di tutta l'umanità […] Chi diventerà leader in questo ambito governerà il mondo […]. La presenza di una posizione di monopolio tecnologico in questo settore è una situazione non desiderabile" (Santagata & Melegari, 2017).

Dal canto suo, il 24 settembre 2024, l'allora Presidente degli Stati Uniti Joe Biden, rivolgendosi all'Assemblea Generale delle Nazioni Unite, ha affermato che "la IA cambierà il nostro modo di vivere, di lavorare e di combattere. [Essa] potrebbe dare il via a un progresso scientifico mai visto prima". Tuttavia, Biden ha anche avvertito che "la IA porta con sé rischi profondi: dai *deepfake* alla disinformazione, fino a nuovi agenti patogeni e armi biologiche. A mano a mano che la IA diventa più potente, [essa] deve diventare più responsabile verso i nostri bisogni e valori collettivi […] non deve fornire ai dittatori strumenti più potenti per incatenare lo spirito umano" (CNN, 2024).

Quanto al suo successore, Donald Trump ha annunciato il lancio di *Stargate*[1], definendolo "il più grande progetto infrastrutturale per la IA mai realizzato negli Stati Uniti" (Roll Call, 2025). Successivamente il 10 luglio 2025, Trump ha presentato *Winning the Race. America's AI Action Plan*. Il piano di azione, distinto in tre pilastri, presenta una sezione per "promuovere l'adozione della IA all'interno del Dipartimento della Difesa". Il documento afferma che "l'Intelligenza Artificiale ha il potenziale per trasformare sia le operazioni belliche sia quelle di back-office del Dipartimento della Difesa. Gli Stati Uniti devono adottare in modo aggressivo l'Intelligenza Artificiale all'interno delle loro Forze Armate se vogliono mantenere la loro preminenza militare" (The White House, 2025a). Per "garantire che gli Stati Uniti diventino il leader indiscusso nella tecnologia dell'Intelligenza Artificiale", la funzionaria Lynne Parker ha definito il piano "il primo passo per garantire [...] il predominio americano in questo settore" (The White House, 2025b).

Dal canto suo il Presidente cinese Xi Jinping, parlando davanti al Politburo ha dichiarato "dobbiamo riconoscere le nostre lacune e raddoppiare gli sforzi per promuovere l'innovazione tecnologica, lo sviluppo industriale e le applicazioni potenziate della IA. È essenziale rafforzare la ricerca di base, concentrandosi sul dominio delle tecnologie come i chip avanzati e i software, costruendo un sistema di IA indipendente, controllabile e collaborativo" (Pomfret & Zhen, 2025). Già nell'ottobre 2017, durante il XIX Congresso del Partito Comunista Cinese il Presidente cinese Xi Jinping ha esortato "l'Esercito Popolare di Liberazione ad accelerare lo sviluppo dell'intelligentizzazione militare" (Sterling, 2020).

Passando al contesto europeo, il 9 febbraio 2025, il Presidente francese Emmanuel Macron ha espresso preoccupazione per il ritardo del Vecchio Continente nella corsa alla IA, dichiarando che "oggi non siamo competitivi. Siamo in ritardo" in quanto l'Europa "ha bisogno di un'agenda sulla IA, perché è necessario colmare il gap con gli Stati Uniti e la Cina". Il Presidente francese ha concluso: "combatterò per la IA. Combatterò per maggiori risposte [nell'ambito] della difesa e della sicurezza europea" (Altman & Quest, 2025).

A seguito delle dichiarazioni di Macron, l'11 febbraio 2025, l'Eliseo ha lanciato la *Dichiarazione di Parigi sul mantenimento del Controllo Umano nei Sistemi d'Arma abilitati alla IA* in cui i firmatari (27 Paesi del mondo, tra cui Germania, Italia e Spagna) hanno riconosciuto "la IA come tecnologia dotata di un potenziale straordinario per trasformare ogni aspetto delle questioni militari ma che, al contempo, solleva grandi sfide e preoccupazioni". A tale fine, gli Stati firmatari si sono impegnati a "sviluppare, implementare e utilizzare le capacità della IA nel dominio militare in modo responsabile, in conformità con il diritto internazionale e in maniera da non compromettere la pace, la sicurezza e la stabilità internazionale". La dichiarazione sottolinea che "la responsabilità e la responsabilizzazione [delle operazioni militari] non possono mai essere trasferite alle

[1] Stargate LLC, fondata da OpenAI, SoftBank, Oracle e MGX, è una *joint venture* americana finalizzata alla creazione di infrastrutture di IA.

macchine, garantendo che l'essere umano rimanga responsabile dell'uso della IA. In linea con questo principio, non autorizzeremo le decisioni di vita o di morte prese da un sistema d'arma autonomo in grado di operare completamente al di fuori del controllo umano o di una catena di comando responsabile". I firmatari della dichiarazione sono concordi nell'affermare che "l'attuale Diritto Internazionale Umanitario si applica pienamente a tutti i sistemi d'arma, compresi quelli dotati di sistemi di IA" (Élysée, 2025).

Il 13 gennaio 2025, il Primo Ministro britannico Keir Starmer ha annunciato l'*AI Opportunities Action Plan*, in vista di una IA che "guiderà un cambiamento straordinario nel Paese. Dall'insegnamento personalizzato al supporto alle piccole imprese […] L'industria della IA ha bisogno di un governo che sia dalla sua parte non uno che resti a guardare mentre le opportunità gli scivolano tra le dita. E, in un momento di feroce competizione, non possiamo restare fermi. Dobbiamo muoverci rapidamente e agire per vincere la corsa globale [alla IA]" (GOVUK, 2025).

Infine, il 13 giugno 2025, il Cancelliere tedesco Friedrich Merz incontrando i vertici dell'azienda NVIDIA ha dichiarato che "gli investimenti in infrastrutture strategiche di IA sono fondamentali per la forza innovativa del nostro Paese" (BPA, 2025).

Anche in Italia, le istituzioni hanno ribadito l'importanza strategica delle tecnologie emergenti nell'ambito della difesa. In una dichiarazione letta in occasione del CyberSec2024 – tenutosi a Roma – il Ministro della Difesa, Guido Crosetto aveva già affermato che la difesa italiana "sta potenziando la digitalizzazione e aggiornando i suoi modelli operativi per garantire la sicurezza al massimo grado […] Questo include l'adozione di tecnologie all'avanguardia come l'Intelligenza Artificiale, i servizi cloud evoluti, le info-strutture spaziali, fino ai nuovi standard di cifratura a protezione delle informazioni. Investire nella formazione di professionalità con specifiche competenze nel settore cyber è – oggi più che mai – essenziale" (ANSA, 2024).

Un anno dopo, il 20 giugno 2025, intervenendo al convegno organizzato dall'università di Padova sul tema *La difesa nazionale e la pace, fra incertezze UE ed egemonia USA,* Crosetto ha ulteriormente sottolineato come "l'Intelligenza Artificiale e il *quantum computing* sono sempre più strumenti di potere. […] Preservare il nostro patrimonio di idee, conoscenza, cultura, valori, identità è oggi una missione di sicurezza nazionale" (Ministero della Difesa, 2025).

5.2. Le posizioni diplomatiche per una normativa sulla IA militare: il Diritto Internazionale Umanitario e la Convenzione su Certe Armi Convenzionali

Il *DIU* regola le leggi della guerra e protegge i civili, limitando l'uso di armi e tattiche che potrebbero causare danni eccessivi. Esso è codificato nelle Convenzioni di Ginevra del 1949 e nei Protocolli Aggiuntivi del 1977 e si basa su tre principi fondamentali che, se correttamente applicati, prevengono l'uso indiscriminato delle armi. I principi chiave del *DIU* sono:

- Principio di distinzione: i belligeranti devono distinguere tra combattenti e civili, nonché tra obiettivi militari e non. In questo quadro, le armi autonome, operando senza costante supervisione umana, rischiano di esporre i civili a gravi pericoli;
- Principio di proporzionalità: un attacco contro un obiettivo militare non deve provocare danni collaterali sproporzionati. Con particolare riferimento alla capacità delle armi autonome di valutare il contesto in tempo reale, essa resta incerta aumentando il rischio di decisioni inadeguate;
- Principio di necessità: le operazioni militari devono essere essenziali per ottenere un vantaggio, senza eccessi. In questo senso, l'uso di armi autonome solleva dubbi su come questo principio venga applicato senza un adeguato controllo umano.

Come non è emersa una definizione comunemente condivisa degli *AWS* non esiste un trattato internazionale vincolante che ne disciplini in modo specifico la ricerca, lo sviluppo e l'impiego. Oltre alle Convenzioni di Ginevra del 1949 e ai Protocolli Aggiuntivi del 1977, il quadro di riferimento generale è fornito dalla *Convenzione su Certe Armi Convenzionali* (*CCW*). Adottata a Ginevra nel 1980 ed entrata in vigore nel 1983, essa mira a limitare o vietare, in conformità con il *DIU*, l'uso di armi particolarmente dannose o che potrebbero colpire indiscriminatamente la popolazione civile. Essa comprende una parte generale e cinque Protocolli che impongono divieti o restrizioni sull'uso di specifiche armi: armi che producono frammenti non rilevabili dai raggi X (Protocollo I); mine, trappole e dispositivi (Protocollo II); armi incendiarie (Protocollo III); laser accecanti (Protocollo IV); residuati bellici esplosivi (Protocollo V)[2] (Abramson, 2017).

Attualmente, la Convenzione conta 128 Stati Parte, incluse le principali potenze militari mondiali, e si distingue per la sua flessibilità, consentendo ai Paesi aderenti di ratificare i singoli Protocolli separatamente. L'architettura istituzionale della *CCW* riflette questa flessibilità, prevedendo normative parallele e incontri periodici tra gli Stati Parte, inclusi incontri annuali e conferenze di revisione quinquennali per valutare l'attuazione della Convenzione. Inizialmente applicabile solo ai conflitti internazionali, nel 2001 la *CCW* è stata estesa ai conflitti armati non internazionali, con le modifiche entrate in vigore nel 2004.

In più occasioni pressoché tutti i rappresentanti dei Paesi Parte hanno sottolineato che la Convenzione e i suoi Protocolli sono strumenti fondamentali per il rafforzamento del *DIU*. Grazie alla sua composizione, che unisce competenze diplomatiche, giuridiche e militari, la Convenzione rappresenta la base giuridica ideale per affrontare le sfide attuali e future legate allo sviluppo e all'uso delle armi autonome.

Resta aperto il dibattito su chi debba essere ritenuto responsabile – il produttore, il comandante o lo Stato utilizzatore – creando una zona grigia normativa. Gli *AWS* rendono difficile, se non impossibile, identificare i responsabili a causa della loro opacità e dell'assenza di controllo umano diretto. Questo vuoto di responsabilità compromette gravemente il diritto al rimedio e alla giustizia: le vittime restano senza tutela, e le violazioni del diritto alla vita da parte di queste tecnologie, senza un meccanismo di

[2] Il testo è consultabile alla pagina: https://treaties.unoda.org/t/ccw.

responsabilità, risultano arbitrarie, ovvero chi debba rispondere nel caso in cui un *AWS* compia un attacco errato (ACHPR, 2024). Inoltre, sono stati evidenziati i pregiudizi tecnologici insiti negli algoritmi che guidano queste armi, i quali potrebbero riflettere discriminazioni sistemiche e colpire in modo sproporzionato determinate categorie di persone. Questo rischio è stato particolarmente denunciato dalle organizzazioni umanitarie, che hanno sollevato l'allarme per una possibile "disumanizzazione digitale" della guerra. Delegare a macchine il potere decisionale su operazioni letali solleva, infatti, profonde questioni etiche, eliminando l'intervento umano e aggravando le disuguaglianze nei conflitti asimmetrici, oltre a incentivare una nuova corsa agli armamenti.

Sul piano tecnologico, le armi autonome rappresentano un settore emergente con sfide irrisolte. La loro capacità di operare in ambienti complessi, distinguendo tra civili e combattenti, rimane incerta. Senza garanzie di affidabilità, il rischio di violazioni del *DIU* è elevato. Molti esperti e attivisti continuano a chiedere un trattato vincolante per regolamentare l'uso della IA militare. Senza una normativa internazionale chiara, i singoli Stati potrebbero svilupparle e impiegarle senza controllo, aumentando il rischio di escalation tecnologica e militare.

5.3. Il Gruppo di Esperti Governativi

Nel 2016, le Alte Parti Contraenti della *CCW* hanno deciso di istituire un *Gruppo di Esperti Governativi* (*GEG*) con il mandato di "esplorare e concordare possibili raccomandazioni sulle opzioni relative alle tecnologie emergenti nel campo degli *AWS*". Il suo primo mandato (2016-2023) è stato contrassegnato da diverse difficoltà che hanno impedito l'avanzamento dei negoziati, tra cui le divergenze su come definire gli *AWS* e se regolarle attraverso un trattato vincolante o una *soft law*.

Sebbene il Gruppo abbia svolto un ruolo cruciale nel discorso internazionale sugli *AWS*, raggiungendo una certa convergenza su un possibile "quadro normativo e operativo", ha ricevuto critiche su più fronti: per il suo focus esclusivo sul *DIU*, a discapito di altre questioni quali l'etica e la sicurezza; per una partecipazione più ristretta rispetto ad altri forum; per la sua attenzione esclusiva agli *AWS* quando altre tecnologie pongono problemi simili; e per il meccanismo procedurale basato sul consenso.

Per strutturare il confronto all'interno del *GEG*, uno dei concetti che ha guadagnato attenzione è il cosiddetto approccio "a due livelli". Introdotto per la prima volta da Francia e Germania nel 2021, questo approccio ha l'obiettivo di suddividere gli argomenti di una diatriba che ha rischiato di bloccare tutto: da un lato, discutere il divieto di determinati tipi di *AWS*; dall'altro, elaborare regolamenti per quei sistemi che non sarebbero rientrati in tale divieto.

Resta da vedere se questo approccio avrà successo, poiché potrebbe essere difficile raggiungere un consenso sulla definizione dei tipi di *AWS* soggetti al divieto (Van den Boogaard, 2024). Fortunatamente, alcune indicazioni utili possono essere tratte dai *Principi Guida* del *GEG* (2019), che affermano:

A. Il *DIU* continua ad applicarsi pienamente a tutti i sistemi d'arma, inclusi lo sviluppo e l'uso potenziale degli *AWS*;
B. La responsabilità umana per le decisioni sull'uso dei sistemi d'arma deve essere mantenuta, poiché la responsabilità non può essere trasferita alle macchine. Questo deve essere considerato in tutto il ciclo di vita del sistema d'arma;
C. L'interazione essere umano-macchina, che può assumere varie forme ed essere implementata in diverse fasi del ciclo di vita di un'arma, dovrebbe garantire che l'uso potenziale dei sistemi d'arma basati su tecnologie emergenti nel campo della IA militare sia conforme al diritto internazionale applicabile, in particolare al *DIU*. Nella determinazione della qualità e dell'estensione dell'interazione essere umano-macchina, dovrebbero essere considerati una serie di fattori, inclusi il contesto operativo e le caratteristiche e le capacità del sistema d'arma nel suo complesso;
D. La responsabilità per lo sviluppo, il dispiegamento e l'uso di qualsiasi sistema d'arma emergente nel quadro della *CCW* deve essere garantita in conformità con il diritto internazionale applicabile, anche attraverso il funzionamento di tali sistemi all'interno di una catena di Comando e Controllo umano responsabile;
E. In conformità con gli obblighi internazionali degli Stati nello studio, sviluppo, acquisizione o adozione di una nuova arma, mezzo o metodo di guerra, deve essere determinato se il suo impiego sarebbe, in alcune o tutte le circostanze, vietato dal diritto internazionale;
F. Quando si sviluppano o acquisiscono nuovi sistemi d'arma basati su tecnologie emergenti nel settore degli *AWS*, devono essere considerati la sicurezza fisica, le adeguate salvaguardie non fisiche (inclusa la sicurezza informatica contro l'*hacking* o la falsificazione dei dati), il rischio di acquisizione da parte di gruppi terroristici e il rischio di proliferazione;
G. Le valutazioni del rischio e le misure di mitigazione devono far parte del ciclo di progettazione, sviluppo, collaudo e dispiegamento delle tecnologie emergenti in qualsiasi sistema d'arma;
H. Si dovrebbe prendere in considerazione l'uso di tecnologie emergenti nel settore della IA militare per sostenere la conformità al *DIU* e ad altri obblighi giuridici internazionali applicabili;
I. Nel formulare potenziali misure politiche, le tecnologie emergenti nel settore della IA militare non dovrebbero essere antropomorfizzate;
J. Le discussioni e le eventuali misure politiche adottate nel contesto della *CCW* non dovrebbero ostacolare i progressi o l'accesso agli usi pacifici delle tecnologie autonome intelligenti;
K. La *CCW* offre un quadro adeguato ad affrontare la questione delle tecnologie emergenti nel settore della IA militare nel contesto degli obiettivi e delle finalità della Convenzione, che cerca di trovare un equilibrio tra la necessità militare e le considerazioni umanitarie.

Nel 2023, il mandato del *GEG* è stato rinnovato per un periodo di tre anni, con il compito di esplorare l'elaborazione di uno strumento normativo per regolamentare la IA

militare, senza determinare se dovesse essere vincolante o meno. Il nuovo presidente del *GEG*, l'ambasciatore del Brasile Flávio Soares Damico, ha cercato di stimolare i negoziati, chiedendo agli Stati di fornire definizioni e spiegazioni più precise su cosa costituisce una tecnologia emergente in questo ambito. Nonostante questi sforzi, il consenso internazionale è continuato a mancare, con posizioni contrastanti su se e come regolamentare la IA militare.

Parallelamente, diversi Stati hanno avviato iniziative nazionali e regionali per promuovere la regolamentazione internazionale della IA militare. In particolare, eventi come il *Rio Seminar* in Brasile, il *Berlin Forum* in Germania, e la Conferenza regionale in Costa Rica hanno sostenuto la creazione di un trattato giuridicamente vincolante, con il Comunicato di Belén che chiede un VI Protocollo Aggiuntivo alla *CCW* per limitare la "disumanizzazione della guerra" (Human Rights Watch, 2023). Nel 2024 l'Austria ha organizzato una conferenza a Vienna per promuovere un dibattito più ampio sugli aspetti giuridici, ma la proposta di avviare un nuovo negoziato multilaterale è stata respinta da molti Stati. Il risultato è stato un documento finale privo di norme vincolanti, comunque incoraggiando il coinvolgimento degli Stati nel Rapporto richiesto dal Segretario Generale.

L'ultima sessione del *GEG* si è tenuta dal 3 al 7 marzo 2025, a Ginevra, dove la discussione si è svolta in un clima più aperto rispetto agli incontri precedenti, con un approfondimento sul controllo umano e sulle implicazioni legali e umanitarie della IA militare. Tuttavia, nonostante il confronto costruttivo, sono riemerse profonde divisioni tra gli Stati, rendendo difficile trovare un consenso su questioni chiave come la definizione di queste armi, il loro possibile divieto e il ruolo dell'essere umano nel loro funzionamento.

Uno dei nodi centrali ha riguardato proprio il controllo umano. Forme significative di controllo umano limitano i rischi che i risultati dei sistemi di IA non rispettino l'intento originale, per identificare tempestivamente errori e conseguenze non intenzionali, nonché per garantire un intervento tempestivo o la disattivazione dei sistemi, qualora ciò si renda necessario (Taddeo *et al*., 2021). La maggioranza dei Paesi ha insistito sull'importanza di garantire che l'uso della forza letale resti sempre supervisionato dall'essere umano, ritenendo inaccettabile delegare una decisione di vita o di morte a una macchina. Altri Stati, tradizionali oppositori di limitazioni troppo stringenti in materia (tra essi Corea del Sud, Israele, Russia, Stati Uniti) hanno invece espresso dubbi sulla fattibilità di questa posizione, sostenendo che il concetto stesso di controllo umano sia troppo vago e aperto a interpretazioni soggettive. Alcuni hanno proposto una formula più sfumata, parlando di "controllo e giudizio umano appropriato", nel tentativo di bilanciare la necessità di supervisione con i progressi tecnologici. Tuttavia, non è stato raggiunto un accordo sul livello di intervento umano necessario per garantire il rispetto del diritto internazionale.

Anche la definizione stessa degli *AWS* è stata oggetto di dibattito. Alcuni Stati, tra cui Irlanda, Messico e Norvegia, hanno sottolineato che il concetto di "letale" non dovrebbe essere interpretato in senso stretto, poiché queste armi potrebbero causare ferite gravi o distruzione senza necessariamente uccidere. Al contrario, Israele, Russia, Singapore e

Stati Uniti hanno difeso una definizione più ristretta, sostenendo che un'arma letale deve essere progettata per uccidere. Questa distinzione ha implicazioni concrete: una definizione troppo rigida potrebbe escludere dalla regolamentazione alcuni sistemi autonomi altamente pericolosi, creando un vuoto normativo difficile da colmare.

Secondo l'United Nations Institute for Disarmament Research (UNIDIR), il concetto di "autonomia" non dovrebbe essere inteso come una qualità assoluta o binaria, ma piuttosto come un grado di libertà che un sistema ha nel raggiungere i propri obiettivi. Questo concetto non va confuso con quello di "intelligenza", che riguarda invece la capacità di un sistema di individuare autonomamente la modalità migliore per ottenere un risultato.

Nel corso degli anni, l'UNIDIR ha osservato come si siano sviluppati tre approcci principali alla definizione degli *AWS*: uno incentrato sugli aspetti tecnici dei sistemi stessi, uno sul ruolo dell'essere umano nel ciclo operativo e uno basato sulle funzioni specifiche che il sistema è in grado di eseguire in modo autonomo. Queste prospettive non sono necessariamente in contrasto tra loro, e anzi potrebbero essere integrate per arrivare a una visione più completa e coerente del concetto di autonomia (Spazian *et al.*, 2021).

Le preoccupazioni umanitarie hanno esercitato un ruolo centrale nel dibattito. Molti Stati hanno sottolineato il rischio che queste armi possano colpire indiscriminatamente o causare danni collaterali inaccettabili. Il tema del divieto degli *AWS* ha visto posizioni fortemente contrastanti. Il Comitato Internazionale della Croce Rossa (CICR) ha chiesto l'eliminazione totale delle armi autonome militari, sostenendo che il loro impiego non è in grado di rispettare il *DIU*. La Campagna internazionale *Stop Killer Robots (SKR)*, che raccoglie oltre 270 organizzazioni in più di 70 Paesi, ha ribadito la necessità di impedire alle macchine di decidere autonomamente sulla vita o sulla morte degli esseri umani (SKR, 2025). Viceversa, Israele, Russia, Stati Uniti si sono opposti fermamente. Ad esempio, gli Stati Uniti hanno sostenuto che "le funzioni autonome e i sistemi d'arma possono essere utilizzati per salvare vite e rafforzare la protezione dei civili. Non dobbiamo privilegiare le armi stupide del passato rispetto a quelle intelligenti del futuro" (Reaching Critical Will, 2025, p. 2).

Un altro punto di scontro ha riguardato la regolamentazione dello sviluppo degli *AWS*. Alcuni attori come Norvegia, Brasile e il CICR, hanno sostenuto che le restrizioni non dovrebbero limitarsi solo all'uso di queste armi, ma intervenire già nella fase di progettazione, per evitare la creazione di tecnologie intrinsecamente pericolose. Altri Paesi, come Francia, Israele e Singapore, hanno invece espresso il timore che una regolamentazione così ampia possa frenare l'innovazione militare e limitare inutilmente lo sviluppo tecnologico.

Il confronto è proseguito nel 2025, con le consultazioni informali tenutesi il 12 e 13 maggio a New York e l'incontro ufficiale previsto dal 1 al 5 settembre a Ginevra. Ad oggi, il dibattito sugli *AWS* ha permesso di chiarire alcune posizioni, ma restano divergenze significative. C'è consenso sulla necessità di una regolamentazione, ma non sulle modalità per attuarla. La definizione stessa degli *AWS* è ancora oggetto di discussione e il rischio di creare vuoti normativi è concreto. Il controllo umano resta un tema critico,

con visioni opposte tra chi lo vuole forte e chi ritiene che possa essere ridotto al minimo. La possibilità di vietare gli *AWS* è fortemente osteggiata dai principali attori militari, mentre non c'è accordo se limitare solo l'uso di queste armi o intervenire già nella fase di sviluppo.

I prossimi mesi saranno decisivi per capire se sarà possibile raggiungere un accordo internazionale su queste tecnologie, che potrebbero cambiare per sempre il volto della guerra. Nel frattempo, l'assenza di regole chiare lascia aperta la possibilità che lo sviluppo delle armi autonome continui senza limiti, con conseguenze difficili da prevedere per la sicurezza globale.

5.4. Il voto ONU sulla IA militare (2 dicembre 2024)

Dal 2018, il Segretario Generale delle Nazioni Unite, António Guterres, sostiene che gli *AWS* sono politicamente inaccettabili e moralmente ripugnanti, invocandone il divieto secondo il diritto internazionale. Nel suo *Nuovo Programma per la Pace* del 2023, ha ribadito questa posizione, raccomandando che gli Stati concludano, entro il 2026, "uno strumento giuridicamente vincolante per vietare i *Sistemi d'Arma Letali Autonomi* che operano senza controllo o supervisione umana e che non possono essere utilizzati nel rispetto del *DIU*, e per regolare tutte le altre tipologie di *Sistemi d'Arma Autonomi*" (UNSG, 2023, p. 27). Ha inoltre osservato che, in assenza di regolamentazioni multilaterali specifiche, la progettazione, lo sviluppo e l'uso di tali sistemi sollevano preoccupazioni umanitarie, giuridiche, di sicurezza ed etiche, e costituiscono una minaccia diretta ai diritti umani e alle libertà fondamentali (UNGA, 2023; UNODA 2023).

Il 2 dicembre 2024, l'Assemblea Generale delle Nazioni Unite ha approvato con ampia maggioranza la risoluzione 79/L.77 sugli *AWS*, con 166 voti a favore, 3 contrari (Bielorussia, Corea del Nord e Russia) e 15 astensioni (Arabia Saudita, Cina, Estonia, Fiji, India, Iran, Israele, Lettonia, Lituania, Nicaragua, Polonia, Romania, Siria, Turchia e Ucraina). Presentata da un gruppo di Stati, tra i quali Austria, Belgio, Brasile, Messico e Svizzera, la risoluzione mira a rafforzare il controllo internazionale su queste tecnologie emergenti e a promuovere un dibattito più strutturato sulla loro regolamentazione.

Pur riconoscendo che strumenti giuridici esistenti – come il *DIU*, la Dichiarazione Universale dei Diritti Umani e la Carta delle Nazioni Unite – si applicano anche alle armi autonome, il documento evidenzia la necessità di normative più specifiche per affrontare le sfide poste dalla IA in ambito militare.

Tra i principali rischi associati all'uso della IA militare, la risoluzione evidenzia:

- Il pericolo di una corsa agli armamenti, che potrebbe acuire le tensioni internazionali e minare la stabilità globale;
- L'abbassamento della soglia per l'uso della forza militare, con potenziali ripercussioni sulla sicurezza internazionale;
- Le implicazioni etiche e umanitarie, legate alla mancanza di controllo umano diretto nelle decisioni di vita o di morte;

- La proliferazione incontrollata, con il rischio che queste tecnologie possano essere utilizzate da attori non statali con fini destabilizzanti.

Per affrontare queste criticità, la risoluzione propone il suddetto approccio su due livelli: da un lato, il divieto di armi che possano non rispettare il *DIU*; dall'altro, una regolamentazione chiara per garantire che il controllo umano rimanga essenziale nelle decisioni sull'uso della forza. Inoltre, prevede il proseguimento dei lavori del *GEG* nell'ambito della *CCW* e la convocazione di consultazioni informali nel 2025 per ampliare il dibattito con la partecipazione di esperti, Stati membri e rappresentanti della società civile (UNGA, 2024a).

La soddisfazione per il voto della risoluzione non ha impedito la denuncia di alcune criticità che permangono. Il documento sottolinea l'urgenza di un trattato internazionale, ma non introduce obblighi immediati per gli Stati. Senza un meccanismo di *enforcement* efficace, il rischio è che le discussioni si prolunghino senza risultati concreti.

Un altro aspetto critico riguarda l'ambiguità nel trattare il ruolo delle tecnologie emergenti. Sebbene il testo riconosca il potenziale positivo della IA, ad esempio nella protezione dei civili, non chiarisce in che modo tali tecnologie possano essere regolamentate per prevenire utilizzi impropri. Anche il coinvolgimento della società civile e della comunità scientifica, sebbene menzionato, non viene accompagnato da strumenti concreti per garantirne una partecipazione effettiva nelle decisioni politiche.

La risoluzione arriva in un momento di crescente preoccupazione per l'impiego delle armi autonome nei conflitti recenti, come in Ucraina e a Gaza. Da oltre un decennio, il *GEG* della *CCW* lavora su questo tema, ma i progressi sono stati lenti a causa del vincolo rappresentato dal consenso unanime tra gli Stati membri, il cui risultato è stato spesso di bloccare decisioni cruciali (Blanchard *et al.*, 2025). La nuova risoluzione, invece, cerca di accelerare il percorso verso un trattato internazionale, invitando a un confronto più aperto e strutturato sulle conseguenze dell'uso degli *AWS*.

Dal punto di vista legale, il dibattito si concentra sulla compatibilità della IA militare con il *DIU*, che impone di distinguere tra civili e combattenti, rispettare il principio di proporzionalità e garantire un controllo umano sufficiente. Anche il diritto penale internazionale potrebbe essere applicato per sanzionare l'uso illecito di queste armi, mentre trattati già esistenti, come la Convenzione di Ginevra, offrono una base normativa, sebbene non specificamente adattata alle tecnologie emergenti.

Le proposte attualmente in discussione includono il divieto di armi autonome che prendano di mira direttamente gli esseri umani, le restrizioni su altri sistemi autonomi, l'obbligo di una supervisione umana adeguata, oltre a meccanismi di autodistruzione o disattivazione in caso di malfunzionamento.

Come sopra menzionato, un elemento chiave della risoluzione è stata la convocazione di consultazioni informali, aperte non solo agli Stati membri, ma anche alla società civile, alla comunità scientifica e all'industria. Svoltesi a New York il 12 e 13 maggio 2025, le consultazioni si sono concluse con alcune riflessioni di Adedeji Ebo – Direttore e Vice Alto Rappresentante per gli Affari del Disarmo – che ha sottolineato l'elevato livello di partecipazione da parte degli Stati e della società civile. In totale, sono intervenuti 39

Stati, l'Unione Europea e 15 organizzazioni civili, a dimostrazione della crescente attenzione e del senso di urgenza che circondano il tema.

È emerso un forte sostegno verso il *GEG* come sede principale per far progredire il dibattito sugli *AWS*. Molte delegazioni hanno richiamato la necessità di evitare processi paralleli e hanno evidenziato l'importanza che queste consultazioni si integrino con il lavoro già avviato a Ginevra. In particolare, si è discusso della centralità del controllo umano sull'uso della forza, della proposta di un approccio normativo su due livelli e della crescente richiesta di un accordo giuridicamente vincolante.

Numerose delegazioni hanno sollevato preoccupazioni di ordine legale, rilevando che l'autonomia delle armi potrebbe diluire le responsabilità giuridiche e compromettere il diritto a un ricorso effettivo. Sul piano dei diritti umani, si è rimarcato il rischio che gli *AWS* non siano in grado di rispettare i principi fondamentali dell'umanità, con implicazioni gravi per la protezione dei civili, in particolare per le persone *hors de combat* e per i combattenti che si arrendono.

È stato sottolineato anche il rischio etico e tecnico derivante da pregiudizi algoritmici legati a razza, genere e abilità, nonché da possibili malfunzionamenti. Molti interventi hanno ribadito l'inaccettabilità morale di delegare a una macchina la decisione sull'uso della forza, richiamando i principi della dignità umana e della coscienza pubblica, come stabilito dalla clausola Martens.

Infine, si è discusso delle ricadute sul piano della sicurezza globale, tra cui l'abbassamento della soglia per il ricorso alle armi, il rischio di escalation, e il pericolo che queste tecnologie vengano integrate con armi di distruzione di massa o diffuse a soggetti non statali. Sono stati inoltre menzionati gli effetti collaterali su infrastrutture, ambiente e popolazioni civili.

Nel ringraziare i partecipanti e gli organizzatori, Ebo ha ribadito la necessità di agire con decisione per rispondere all'appello del Segretario Generale delle Nazioni Unite, che chiede l'adozione di uno strumento giuridicamente vincolante sugli *AWS* entro il 2026. Ha infine assicurato che l'Ufficio per gli Affari del Disarmo continuerà a sostenere questi sforzi, sia a New York sia a Ginevra (Ebo, 2025).

Nonostante il voto ONU sia stato accolto positivamente, molte organizzazioni della società civile lo considerano insufficiente, poiché il testo adottato non riflette l'urgenza di avviare negoziati su uno strumento giuridicamente vincolante (Rete Italiana Pace Disarmo, 2024). La pressione della comunità internazionale per una regolamentazione più chiara è aumentata, anche a causa delle tensioni globali e della competizione tra potenze come Stati Uniti, Russia e Israele, che stanno investendo massicciamente nello sviluppo di armi autonome avanzate.

In definitiva, sebbene il diritto internazionale imponga già alcuni limiti all'uso della IA militare, la mancanza di un quadro normativo chiaro lascia spazio a interpretazioni discordanti e apre il rischio che queste tecnologie vengano utilizzate in modo indiscriminato. La risoluzione ONU del 2024 rappresenta un passo importante verso un accordo globale per garantire un controllo efficace su queste armi. La vera sfida sarà convincere tutti gli Stati, soprattutto quelli con grandi programmi di sviluppo militare, a

impegnarsi per un trattato vincolante che metta al centro la sicurezza globale e le implicazioni etiche della IA in ambito bellico.

5.5. Posizioni nazionali sull'interdizione delle armi autonome letali e non

Nel contesto dei negoziati internazionali sulle armi autonome, gli Stati membri delle Nazioni Unite hanno espresso una varietà di posizioni. Sintetizzandole, emergono tre principali orientamenti nazionali riguardo ai *Sistemi d'Arma Autonomi* e non (Automated Decision Research, 2025).

In primo luogo, vi sono alcuni Stati – da noi definiti *Liberalizzatori* – i quali ritengono che l'applicazione del diritto internazionale esistente sia sufficiente per regolamentare gli *AWS* e quindi non vi sia necessità di specifici divieti. Secondo i più espliciti, la IA nei sistemi di difesa e nelle operazioni militari costituisce una forma di innovazione necessaria per mantenere la superiorità tecnologica[3].

Sul fronte opposto si situano altri Stati che propugnano una posizione "proibizionista", chiedendo l'adozione di un trattato che vieti categoricamente gli *AWS*. Questi sottolineano il rischio che l'uso indiscriminato degli *AWS* possa sfuggire a qualsiasi forma di responsabilità legale, compromettendo la sicurezza globale e alimentando conflitti in modo incontrollato. Inoltre, sollevano preoccupazioni etiche, sostenendo che l'uso della forza letale da parte di una macchina priva di giudizio umano è moralmente inaccettabile, riducendo la capacità di una società di esercitare il controllo sul proprio destino e le sue decisioni più critiche in contesti di guerra.

Infine, una terza posizione è quella degli Stati che propongono un approccio differenziato, ovvero l'introduzione di un trattato che proibisca alcuni utilizzi degli *AWS*, consentendone altri sotto regolamentazioni specifiche. In questo modo, si cerca di stabilire un controllo più preciso, limitando l'uso delle armi autonome in determinate situazioni più circoscritte e monitorate (Perrin & Zamani, 2025).

5.5.1. I "Liberalizzatori": contrari a divieti legalmente vincolanti

Il primo gruppo di Stati è contrario all'adozione di una proibizione formale degli *AWS* ritenendo che:

- Il *DIU* vigente è già sufficiente;
- Un divieto vincolante limiterebbe lo sviluppo tecnologico militare;
- Siano da preferire principi volontari e dichiarazioni politiche (v. tab. 5.1).

[3] Nella *constituency* tecnologica di riferimento si fa osservare che la IA aumenta l'efficacia e l'efficienza delle missioni grazie all'impiego di sensori avanzati e tecniche sofisticate di analisi di grandi quantità di dati. Questo permetterebbe, ad esempio, di implementare sistemi di manutenzione predittiva essenziali per gli assetti mobili. Inoltre, gli algoritmi della IA migliorano la capacità decisionale in scenari complessi, supportando i sistemi di Comando e Controllo nell'ottimizzazione delle risorse e nell'adozione di strategie. Le sue potenzialità si estendono alla pianificazione strategica e alla simulazione di scenari operativi, facilitando l'addestramento delle truppe e la logistica militare (AFCEA Roma, 2023).

Tab. 5.1. AWS: Stati "Liberalizzatori"

Stato	Posizione
Stati Uniti	Si oppone a nuovi divieti vincolanti. Preferisce principi volontari.
Russia	Si oppone a limitazioni legali. Vede gli *AWS* come sviluppo naturale.
Israele	Si oppone a una nuova regolazione legale vincolante.
Cina	Pur riconoscendo la necessità di regolamentazione, non sostiene un divieto vincolante. Sottolinea l'importanza di un uso responsabile delle tecnologie emergenti.
India	Ritiene prematuro parlare di proibizione. Preferisce chiarimenti sull'applicazione del diritto esistente prima di negoziare nuovi strumenti legali.

Elaborazione Archivio Disarmo su UNGA, 2024b

La posizione liberalizzatrice è sostenuta con particolare vigore da Paesi come Israele, Stati Uniti e Russia, seguiti da altri come l'India e la Cina i quali, pur non opponendosi totalmente, assumono una posizione più sfumata. Nel gruppo dei "liberalizzatori", la Federazione Russa ritiene che "attualmente non vi siano motivi convincenti per imporre nuove limitazioni o restrizioni sugli *AWS*, né per aggiornare o adattare il *DIU* per affrontare tali armi" (UNGA, 2024b, p. 95). Secondo la Russia, gli Stati dovrebbero affrontare queste questioni individualmente, come stabilito dall'Articolo 36 del Protocollo Aggiuntivo I della Convenzione di Ginevra del 1949, che obbliga gli Stati a determinare se l'uso di una nuova arma violerebbe il diritto internazionale applicabile. La Russia sottolinea che l'Articolo 36 non impone requisiti specifici su come condurre le revisioni legali delle nuove armi, lasciando la questione alla discrezione degli Stati.

Gli Stati Uniti, pur sostenendo una visione simile circa la discrezionalità dei singoli Stati, pongono l'accento sull'importanza di implementare le strategie esistenti nel quadro del *DIU*, sottolineando che quest'ultimo "non proibisce l'uso dell'autonomia nei sistemi d'arma". Gli Stati Uniti respingono il concetto di *Controllo Umano Significativo* (*CUS*), sostenendo che non esiste un livello fisso di giudizio umano che deve essere applicato a ogni contesto, suggerendo che il controllo umano dovrebbe variare in base alle politiche nazionali.

Anche la Russia si oppone ai termini "vaghi" come il *CUS*, sostenendo che esso dovrebbe essere a discrezione degli Stati e che non sia necessario un controllo diretto, ma piuttosto misure pratiche per garantire l'affidabilità e la tolleranza agli errori dei sistemi. Sia la Russia sia gli Stati Uniti si dichiarano favorevoli a mantenere il controllo umano sui sistemi autonomi, ma con approcci distinti. La Russia ritiene che il controllo umano debba essere flessibile e consentire l'intervento umano quando necessario, ma non obbligatorio in ogni circostanza, mentre gli Stati Uniti sottolineano che il controllo umano dovrebbe essere applicato in modo pratico per garantire che le decisioni delle macchine siano coerenti con le intenzioni dei comandanti, adottando misure concrete in tutte le fasi – progettazione, sviluppo e impiego – per riflettere tali intenzioni, ridurre i rischi di azioni

non volute e permettere un adeguato esercizio del giudizio umano nell'uso della forza (UNGA, 2024b, p. 115).

Nel contesto di una possibile regolamentazione internazionale, entrambi i Paesi concordano sull'importanza di un dibattito multilaterale. Gli Stati Uniti, in particolare, apprezzano il lavoro svolto nel *GEG* della *CCW*, ritenendo che questa sia la migliore istanza nella quale sviluppare uno strumento condiviso per affrontare le tecnologie emergenti. La Russia è d'accordo con questa visione e sostiene che il *GEG* è il contesto giusto per bilanciare le preoccupazioni umanitarie con gli interessi difensivi degli Stati, evitando il trasferimento del dibattito su altre piattaforme internazionali, comprese le Nazioni Unite (UNGA, 2024b, p. 94).

Israele, pur riconoscendo i rischi associati agli *AWS*, sostiene che "i sistemi d'arma basati su tecnologie emergenti nel campo degli *AWS* possono rispondere sia a necessità militari sia a considerazioni umanitarie, e possono essere impiegati per garantire il rispetto del *DIU*". Israele avverte che iniziative esterne alla *CCW* potrebbero frammentare gli approcci esistenti e sottolinea che "nella discussione sugli *AWS*, l'attenzione debba essere rivolta all'applicazione del *DIU* esistente, tenendo conto del contesto operativo. È problematico discutere della legalità degli *AWS* focalizzandosi solo sulle capacità dell'arma" (UNGA, 2024b, pp. 61-62).

Paesi come la Cina e l'India, pur riconoscendo il potenziale rivoluzionario degli *AWS*, sottolineano la necessità di un approccio cauto, che bilanci i benefici tecnologici con le esigenze umanitarie. La Cina, in particolare, si mostra favorevole a un divieto parziale degli *AWS*, ritenendo che alcune tecnologie possano comportare rischi legali ed etici significativi. La Cina promuove inoltre una governance globale della IA in linea con il diritto internazionale, facendo leva sulla necessità di misure di controllo umano.

Nel Sud del mondo, numerosi Paesi, inclusi quelli con minori capacità tecnologiche, esprimono preoccupazioni circa le implicazioni legali ed etiche degli *AWS*, temendo che il loro uso incontrollato possa generare escalation nei conflitti e violazioni dei diritti umani. Tuttavia, alcuni Paesi, come il Pakistan (coinvolto con l'India in un contenzioso che ha avuto nella primavera 2025 una recrudescenza tanto più rischiosa in quanto riguardante due Paesi nucleari), sono impegnati nello sviluppo della IA per non restare indietro ciascuno rispetto al proprio vicino.

5.5.2. I "Proibizionisti": favorevoli a una proibizione totale

A differenza dei *Liberalizzatori* che vi si oppongono, alcuni Stati – definibili come *Proibizionisti* – chiedono esplicitamente la negoziazione urgente di un trattato internazionale legalmente vincolante che proibisca lo sviluppo, il possesso e l'uso di armi autonome legali prive di *CUS*, considerate incompatibili con i principi etici e con il *DIU* (Hoppenbrouwers, 2024). Tali Stati, spesso firmatari del Comunicato di Belén[4] e/o

[4] Il testo è consultabile alla pagina: https://www.rree.go.cr/files/includes/files.php?id=2261&tipo=documentos.

sostenitori della bozza di Protocollo VI alla *CCW*, condividono l'obiettivo di escludere questo tipo di armi dal lecito uso bellico (v. tab. 5.2).

Tab. 5.2. AWS: Stati "Proibizionisti"

Stato	Posizione
Kiribati	Sostiene una proibizione totale tramite trattato legalmente vincolante.
Argentina	Promuove "proibizioni mirate", cioè il divieto specifico di quelle applicazioni degli *AWS* che risultano incompatibili con i principi etici e giuridici, soprattutto quando manca il *CUS* sulle funzioni critiche.
Cile	Sostenitore del Protocollo VI e del divieto totale.
Colombia	Sostenitore del Protocollo VI.
Costa Rica	Chiede una proibizione totale degli *AWS* privi di *CUS*.
Messico	Richiede uno strumento giuridico per proibire gli *AWS* privi di *CUS*.
Panama	Firmatario del comunicato regionale per una proibizione vincolante.
Perù	Firmatario di iniziative per un trattato con proibizioni.
Uruguay	Sostiene un trattato per proibire gli *AWS* privi di *CUS*.
Rep. Dominicana	Sostenitore del comunicato di Belén.
Cuba	Sostiene la negoziazione di uno strumento legalmente vincolante per vietare l'uso degli *AWS* privi di *CUS*.
Serbia	Sostiene il divieto degli *AWS* o la limitazione della loro produzione e un rigoroso controllo sull'impiego. Sottolinea la non conformità di questi sistemi ai principi del *DIU*.

Elaborazione Archivio Disarmo su UNGA, 2024b

Al gruppo di Stati che si distingue per la posizione fermamente contraria, appartengono Paesi come Cuba, che sostiene l'adozione di un protocollo internazionale vincolante volto a vietare l'uso degli *AWS*, promuovendo una regolamentazione rigorosa delle tecnologie autonome. Cuba insiste sul divieto totale degli *AWS* proponendo altresì una normativa severa per le armi semi-autonome, come i droni, al fine di prevenire danni umanitari e garantire che tali tecnologie non sfuggano al controllo umano. Tale posizione trova concorde il parere di altri Stati, come la Serbia e le isole Kiribati (già coinvolti negli anni Sessanta dalle esplosioni nucleari sperimentali degli Stati Uniti), che chiedono un'interdizione totale delle armi autonome, sottolineando l'importanza del *CUS*.

Nel contesto della *CCW*, Cuba riconosce il valore dei dibattiti svolti negli ultimi anni sullo sviluppo e l'uso delle armi autonome, ma ritiene che sia necessario andare oltre la semplice discussione e adottare un trattato che vieti esplicitamente la produzione e l'impiego di tali sistemi, in linea con le preoccupazioni etiche, giuridiche e di sicurezza internazionale. L'Avana sottolinea, infatti, che le armi autonome non sono compatibili con i principi fondamentali del *DIU*, in particolare per quanto riguarda i principi di distinzione tra combattenti e civili e di proporzionalità nell'uso della forza.

Paesi come la Serbia e Kiribati esprimono una preoccupazione particolare per le implicazioni etiche e morali dell'impiego di sistemi autonomi. Tali Stati temono che l'uso di queste armi possa innescare un'escalation di violenze, destabilizzare la sicurezza globale e, a parere del rappresentante della Repubblica di Kiribati, influire negativamente sulle decisioni riguardanti il controllo delle armi nucleari. Questo arcipelago dell'Oceano Pacifico, che ha sperimentato le devastanti conseguenze dei test nucleari, si oppone fermamente all'impiego di sistemi autonomi che potrebbero essere utilizzati per la gestione degli armamenti nucleari. La Serbia, invece, pone l'accento sul rischio di danni collaterali e sulla difficoltà di attribuire le responsabilità per le violazioni che potrebbero derivare dall'uso delle armi autonome. Anche la Costa Rica, Paese da sempre impegnato in favore della pace e della sicurezza globale, ha sollevato preoccupazioni etiche e legali circa l'impiego degli *AWS*. La Costa Rica ha spinto per l'adozione di un trattato internazionale che vieti i sistemi d'arma privi di *CUS*, in quanto tali sistemi non sono compatibili con il *DIU* e potrebbero essere facilmente manipolati da attori non statali, come gruppi terroristici o criminali. Il Paese ha proposto, inoltre, l'inclusione di una prospettiva multidisciplinare nei negoziati, che consideri non solo il *DIU*, ma anche gli aspetti etici, i diritti umani e la sicurezza globale.

5.5.3. I "Dualisti": favorevoli a un approccio a due livelli (divieto + regolazione)

Per conciliare le preoccupazioni etiche con le esigenze politiche, un terzo gruppo di Stati propone l'approccio noto come "dualista". Formalizzato nel 2021 dal Comitato Internazionale della Croce Rossa, "l'approccio a due livelli", mira a trovare un compromesso tra la regolamentazione e il divieto degli *AWS*[5] (CICR, 2021). Tale modello prevede:

- Il divieto assoluto per gli *AWS* che non possano essere utilizzati nel rispetto del *DIU*, in particolare quelli che sono imprevedibili o destinati esplicitamente a colpire esseri umani;
- La regolamentazione rigorosa per tutte le altre forme degli *AWS*, con meccanismi di controllo, supervisione e responsabilità umana (v. tab. 5.3).

[5] Per un approfondimento su ciascuno dei due livelli, v. *supra* par. 1.2.

Tab. 5.3. AWS: Stati "Dualisti"

Stato	Posizione
Germania	Divieto vincolante per alcuni degli *AWS* e regolamenti per altre.
Francia	Divieto per sistemi incontrollabili, regolazione per altri.
Italia	Sostenitrice dell'approccio duale.
Danimarca	Co-firmataria della proposta sul *two-tier approach*.
Norvegia	Divieto per alcune categorie e regolazione delle altre.
Lussemburgo	Divieto degli *AWS* non conformi al *DIU*, regolazione delle altre.
Svezia	Considera il *two-tier approach* come base di consenso.
Canada	Sostiene proibizione e regolazione nel quadro del *GEG*.
Paesi Bassi	Divieto per gli *AWS* non compatibili con il *DIU*, regolazione per le altre.
Bulgaria	Propone il *two-tier approach* come quadro guida.
Austria	Sottolinea la necessità di stabilire principi fondamentali per l'uso delle armi autonome.
Svizzera	Riconosce i benefici e i rischi delle tecnologie autonome, enfatizzando l'urgenza di regolamentare il loro sviluppo per garantire il rispetto del *DIU*.

Elaborazione Archivio Disarmo su UNGA, 2024b

Questo approccio è sostenuto da numerosi Stati che lo vedono come una via per garantire la sicurezza internazionale, mantenendo al contempo il *CUS* sui sistemi d'arma.

L'approccio a due livelli distingue tra i sistemi d'arma che possono essere effettivamente controllati dagli esseri umani e quelli che non offrono garanzie di supervisione. I sistemi che rientrano nella prima categoria sono soggetti alle normative generali del diritto internazionale, mentre quelli che appartengono alla seconda categoria dovrebbero essere vietati. I Paesi Bassi, ad esempio, definiscono "imprevedibilità intrinseca" la capacità di un sistema di modificare autonomamente compiti, obiettivi o regole di ingaggio senza l'approvazione di un operatore umano. In base a questa definizione, un trattato vincolante dovrebbe impedire la progettazione e lo sviluppo di tali sistemi, che non possiedono alcun tipo di controllo umano. Tuttavia, una delle principali sfide in queste discussioni riguarda la definizione precisa di cosa costituisca un "controllo umano" o una "approvazione umana", dato che i confini tra supervisione e autonomia non sono ancora chiaramente delineati.

Da un lato, i sostenitori di questo approccio (tra cui l'Italia), propongono di vietare i sistemi che non rispettano i principi fondamentali del *DIU*, come la capacità di distinguere tra combattenti e civili o di effettuare valutazioni di proporzionalità negli attacchi. Dall'altro lato, ritengono che questi sistemi, seppure autonomi, siano in grado di rispettare tali principi e perciò debbano essere regolamentati anziché vietati. L'Italia, ad esempio, definisce questi sistemi come quelli che, pur avendo capacità autonome, possono essere valutati rispetto alla loro conformità al *DIU* attraverso test adeguati e formazione degli operatori. L'approccio italiano, che propone l'applicazione di un quadro normativo che

distingue tra i sistemi che rispettano e quelli che violano il diritto internazionale, ha ottenuto un ampio supporto da altri Stati, come l'Austria, che ha recentemente sottolineato la necessità di stabilire principi fondamentali per l'uso delle armi autonome. La Svizzera ha espresso una posizione simile, riconoscendo sia i benefici sia i rischi derivanti dall'uso delle tecnologie autonome, e ha sottolineato l'urgenza di regolare lo sviluppo e l'impiego di questi sistemi per garantire il rispetto del diritto internazionale e degli obblighi umanitari (DFAE, 2024).

Le posizioni espresse da questi Stati convergono sull'importanza di stabilire un *CUS* in ogni fase dell'uso della forza, assicurando che le decisioni cruciali, come quelle relative alla vita e alla morte, restino sotto la responsabilità di esseri umani. L'adozione di un trattato vincolante che regoli l'uso degli *AWS*, con particolare attenzione alla trasparenza, alla prevedibilità e alla responsabilità, è quindi vista come una necessità urgente per prevenire la disumanizzazione della guerra e garantire la protezione dei civili.

Le divergenze tra queste tre correnti delineano uno scenario geopolitico complesso. Mentre molti Stati del Sud globale e dell'Europa spingono per un approccio etico e legale forte, le maggiori potenze tecnologiche si mostrano restie a vincoli giuridici. Tuttavia, l'emergere di un consenso parziale sull'approccio a due livelli potrebbe rappresentare un punto di convergenza pragmatico per il futuro negoziale.

5.6. Altre Istituzioni internazionali

Nel complessivo panorama delle posizioni, un significativo rilievo è rivestito anche da altre Istituzioni, entità politiche come l'Unione Europea oppure come la NATO. A questo livello importanti contributi volti a costruire cornici normative alternative o complementari ai negoziati della *CCW* sono stati offerti in tema di regolazione della IA militare.

5.6.1. La governance multilivello: iniziative globali

A fronte dell'assenza di un trattato vincolante a livello globale, dal 2023 si sono affermate due iniziative multilivello di riferimento: il Summit REAIM (*Responsible Artificial Intelligence in the Military Domain*) e la *Dichiarazione Politica sull'Uso Responsabile della IA Militare*.

Il Summit REAIM, avviato da Corea del Sud e Paesi Bassi, si propone come piattaforma aperta al dialogo tra governi, società civile e comunità scientifica. La prima edizione, tenutasi all'Aja, ha visto la partecipazione di duemila delegati provenienti da un centinaio di Paesi. Cinquantasette Stati, perlopiù europei, hanno sottoscritto una dichiarazione congiunta che incoraggia uno sviluppo della IA militare improntato alla responsabilità, alla trasparenza e al coinvolgimento non solo delle autorità politiche, ma anche di accademici, esperti e organizzazioni indipendenti. L'impostazione di REAIM è inclusiva e deliberatamente "dal basso": mira a soluzioni collaborative e adattabili, in grado di tener conto della pluralità degli attori coinvolti. Questo approccio ha permesso

al vertice di attrarre anche Paesi non occidentali, come la Cina e vari membri del Sud globale, tradizionalmente cauti verso iniziative percepite come espressione delle potenze atlantiche.

In un'ottica diversa, più istituzionalizzata e verticale, si colloca invece la *Dichiarazione Politica* promossa dagli Stati Uniti e presentata in occasione del Summit REAIM del 2024. Rivolta esclusivamente agli Stati, la *Dichiarazione* propone un quadro di regole operative e impegni politici finalizzati a garantire un impiego sicuro e responsabile della IA nei sistemi d'arma e nei processi decisionali militari. L'iniziativa, sottoscritta da 58 Paesi – tra cui tutti gli Stati membri dell'UE – punta a rafforzare la trasparenza, la supervisione umana e la responsabilità nelle fasi di progettazione, sperimentazione e impiego operativo (DoS, 2023). La precedente amministrazione Biden aveva sostenuto la *Dichiarazione*, nel tentativo di affermare una leadership normativa americana in questo dominio emergente, contenendo al tempo stesso le ambizioni cinesi e preservando il primato tecnologico statunitense.

Con l'elezione di Donald Trump nel 2025 e il conseguente cambio di orientamento strategico, l'atteggiamento degli Stati Uniti nei confronti della governance multilaterale della IA sembra però essere mutato sensibilmente. L'amministrazione Trump ha mostrato fin dai primi mesi un interesse più marcato per l'autonomia strategica e la competizione tecnologica, ridimensionando l'impegno cooperativo in favore di una logica di vantaggio nazionale. Questo cambio di passo ha ridotto il peso politico della *Dichiarazione*, che rischia ora di perdere efficacia come strumento di *soft law*, specie agli occhi dei Paesi del Sud globale, già inclini a percepirla come un'iniziativa occidentale calata dall'alto.

Pur muovendosi su piani distinti, REAIM e la *Dichiarazione Politica* possono essere letti in chiave complementare. Il Summit può offrire uno spazio di confronto più informale, inclusivo e multilivello, in grado di legittimare e rafforzare l'accettabilità delle linee guida più tecniche e stringenti contenute nella *Dichiarazione*. Tuttavia, la fragilità dell'attuale architettura politica globale rappresenta un ostacolo non trascurabile. Il disallineamento crescente tra Stati Uniti ed UE, l'assenza strutturale di attori chiave come Russia e Cina e la persistente sfiducia di molti Paesi del Sud globale verso meccanismi percepiti come unilaterali rendono difficile l'affermazione di un consenso effettivo.

In questo quadro instabile, l'UE potrebbe esercitare un ruolo importante.

Sebbene priva delle capacità industriali e strategiche delle grandi potenze, l'UE dispone di una credibilità normativa che le consentirebbe di porsi al centro di uno sforzo di convergenza. Promuovere standard condivisi, favorire il dialogo tra Nord e Sud del mondo e mantenere aperti i canali tra REAIM e la *Dichiarazione* rappresentano ambiti in cui l'Europa può investire, facendo leva sul suo capitale politico e diplomatico per garantire legittimità, inclusività e stabilità alla futura governance della IA militare.

5.6.2. L'Unione Europea: la IA tra centralità dei diritti umani e regolazione delle applicazioni civili

Nell'Unione Europea, il Parlamento Europeo ha più volte sostenuto la necessità di un trattato internazionale che vieti l'uso degli *AWS*. Al centro della sua posizione vi è il principio del *CUS*, che afferma la necessità di un intervento umano nelle decisioni sull'uso della forza letale, per garantire responsabilità e rispetto del diritto internazionale, in particolare del *DIU*.

L'UE riconosce che tecnologie emergenti come la IA stanno trasformando il contesto bellico e la sicurezza globale. Per questo promuove un impegno collettivo nella definizione di norme e regolamenti che assicurino un impiego responsabile di tali tecnologie, bilanciando le opportunità offerte con i rischi potenziali, anche in ambito militare.

A livello internazionale, l'UE incoraggia il dialogo e la cooperazione per condividere le *best practices*, sviluppare una comprensione comune delle implicazioni etiche e legali della IA e sostenere iniziative regionali e globali sul tema. Particolare attenzione è riservata alla gestione dei dati, fondamentali per il funzionamento dei sistemi intelligenti, e alla prevenzione dei *bias* algoritmici, compresi quelli di genere, che potrebbero amplificare discriminazioni esistenti.

Secondo il Consiglio dell'UE, l'Unione considera la *CCW* un quadro adatto per affrontare la questione degli *AWS* e continua a partecipare attivamente ai lavori del *GEG*, con l'obiettivo di rafforzare il rispetto del diritto internazionale, anche sul piano etico (Consiglio dell'Unione Europea, 2021).

L'approccio europeo si articola su due livelli: da un lato, gli Stati dovrebbero evitare lo sviluppo e l'uso di sistemi d'arma non compatibili con il diritto internazionale; dall'altro, per i sistemi con componenti autonome è necessaria una regolamentazione rigorosa che ne garantisca la conformità legale ed etica.

L'UE valorizza il contributo delle conferenze e delle iniziative internazionali in materia e sottolinea l'importanza di integrare una prospettiva di genere nelle politiche sulle tecnologie emergenti, riconoscendo il legame tra parità di genere e innovazione responsabile.

Facendo un breve riferimento, data la natura solo contestuale dell'oggetto, alla normazione europea, essa è rappresentata dal Regolamento UE 2024/1689, adottato dal Parlamento Europeo e dal Consiglio il 13 giugno 2024, con l'obiettivo di stabilire un quadro giuridico comune per regolamentare l'uso della IA all'interno dell'Unione. Questo regolamento si concentra sull'impiego civile e commerciale della IA, imponendo delle specifiche normative per garantire la sicurezza, la trasparenza e il rispetto dei diritti fondamentali nell'utilizzo di questa tecnologia (Gazzetta Ufficiale dell'Unione Europea, 2024). Tuttavia, sono esenti dal regolamento i sistemi di IA usati per scopi militari, di difesa o di sicurezza nazionale. Questo vale sia per i sistemi gestiti da enti pubblici, come i governi, sia per quelli gestiti da aziende private che forniscono tecnologia alle Forze Armate. La ragione principale di questa esenzione risiede in due aspetti legali chiave.

Il primo riguarda l'Articolo 4, paragrafo 2, del Trattato sull'Unione Europea, che stabilisce che la politica estera e di sicurezza comune è una competenza esclusiva degli Stati membri (Gazzetta Ufficiale dell'Unione Europea, 2012). Ciò significa che ogni Stato ha la responsabilità di regolamentare i propri sistemi di difesa, senza che l'Unione Europea intervenga direttamente su questo fronte. Il secondo aspetto è legato alle politiche di difesa comune dell'UE, che permettono a ciascun Stato membro di decidere in autonomia come utilizzare le proprie tecnologie militari in linea con le proprie esigenze di sicurezza nazionale.

Data l'importanza strategica che riveste lo sviluppo delle tecnologie di IA in ambito militare, è fondamentale che gli Stati europei mobilitino le proprie risorse e capacità al fine di accelerare la definizione di un quadro di governance comune. Pur non essendo in grado di dissuadere in modo definitivo attori come la Russia e altri Stati affini, secondo le fonti UE l'adozione e la sottoscrizione di un insieme articolato di principi condivisi sulla IA bellica da parte del maggior numero possibile di Paesi responsabili contribuirebbe quantomeno ad accrescere la pressione politica sugli Stati "liberalizzatori".

5.6.3. La NATO: un approccio tecnologico e strategico

Per quanto riguarda la NATO, l'organizzazione adotta un approccio "pragmatico" all'uso della IA, riconoscendone il valore strategico ma senza promuovere esplicitamente lo sviluppo degli *AWS*. Nel 2021 ha infatti adottato la *NATO Artificial Intelligence Strategy* una strategia che, pur non prevedendo un divieto specifico, mira a garantire un impiego responsabile della IA in ambito militare, bilanciando vantaggi operativi con considerazioni etiche e legali. Questo documento politico e coordinativo definisce come la NATO intende guidare e regolamentare l'adozione della IA nel settore della difesa (NATO, 2021).

Nel luglio 2024 la strategia NATO sull'Intelligenza Artificiale è stata aggiornata, in risposta all'evoluzione rapida e significativa delle tecnologie di IA, con particolare attenzione allo sviluppo della IA generativa e dei modelli fondazionali. Queste tecnologie, in grado di generare testi complessi, immagini, audio e codice con qualità sempre più vicina a quella umana, impongono alla NATO e agli Alleati una risposta più concreta, che supera il precedente approccio dichiarativo.

Una delle principali novità è proprio l'urgenza di accelerare l'adozione della IA nelle capacità militari alleate, non solo in termini teorici, ma attraverso casi d'uso reali, l'integrazione nei processi di pianificazione della difesa e l'impiego di strumenti come DIANA (*Defence Innovation Accelerator for the North Atlantic*) e il *NATO Innovation Fund* per spingere l'innovazione. Il nuovo documento introduce inoltre una struttura tecnica più solida per garantire che l'adozione della IA sia eseguita in modo responsabile. In questo senso, assume rilievo la creazione di un sistema NATO per il *Testing, Evaluation, Verification & Validation* (TEV&V), utile a testare e certificare sistemi di IA,

verificandone l'affidabilità, la sicurezza e la coerenza con i Principi di Uso Responsabile già definiti.

Un altro aspetto innovativo riguarda la crescente attenzione alle minacce ibride e all'uso malevolo della IA. L'aggiornamento sottolinea i rischi legati alla disinformazione, all'utilizzo della IA per diffondere narrazioni manipolatorie o violenza di genere digitalmente facilitata, e alla potenziale destabilizzazione delle democrazie e delle Forze Armate.

La strategia aggiornata dà anche maggior peso all'interoperabilità tra sistemi di IA degli Alleati e alla standardizzazione, due elementi fondamentali per garantire coerenza e cooperazione in ambito militare. Contestualmente, si rafforza la cooperazione con il settore privato, l'accademia e i fornitori non tradizionali, riconoscendo che la competitività tecnologica si gioca anche fuori dai tradizionali ambiti militari.

Un altro punto centrale riguarda il ruolo strategico dei dati: si riconosce esplicitamente che la IA ha bisogno di dati di qualità per funzionare in modo efficace e imparziale, e che occorre lavorare per proteggere questa risorsa, assicurando al contempo la gestione del *bias* e il rispetto della privacy e della sicurezza.

Infine, il documento sottolinea l'importanza di un adeguamento delle competenze umane, promuovendo lo sviluppo di una forza lavoro *AI-ready*, con investimenti in formazione, aggiornamento e maggiore integrazione di esperti all'interno delle Forze Armate (NATO, 2024).

Secondo fonti ufficiali "l'integrazione della IA nelle Forze Armate degli Alleati è già iniziata e rappresenta una tendenza irreversibile. Perché la NATO mantenga il suo vantaggio tecnologico e la sua superiorità nella difesa collettiva, è essenziale concentrarsi sull'interoperabilità futura ed evitare innovazioni isolate. […] Inoltre, con l'ampliarsi del divario di capacità nazionali nel campo della IA, la NATO deve assicurarsi che i membri meno equipaggiati, o l'Alleanza stessa, siano in grado di rispondere agli avversari che utilizzano tecnologie basate sulla IA e/o sistemi autonomi in un futuro contesto di conflitto caratterizzato da una maggiore rapidità" (Assemblea Parlamentare NATO, 2024, p.11).

A tal fine, la NATO promuove la consapevolezza sull'uso della IA, puntando a un equilibrio tra innovazione e responsabilità. Ritiene necessario aggiornare regolarmente strategie e linee guida per stare al passo con l'evoluzione tecnologica, e incoraggia la collaborazione con il settore privato per sviluppare standard comuni che facilitino l'interoperabilità. In parallelo, sostiene la regolamentazione internazionale della IA, lavorando con organizzazioni come l'UE per norme coerenti tra ambiti civili e militari. Particolare attenzione è dedicata ai rischi etici, come i *bias* algoritmici, per evitare discriminazioni e garantire un uso equo della tecnologia.

La NATO riconosce l'importanza del dialogo con attori globali come Cina e Russia, con l'obiettivo di definire norme condivise sull'uso responsabile della IA e rafforzare la sicurezza internazionale attraverso una governance etica e inclusiva. "Resta da vedere se questi Paesi si considerino realmente parte interessata nella definizione di un insieme globale di regole, anche se, ad esempio, lo scenario di un utilizzo della IA per sviluppare

armi chimiche e biologiche dovrebbe risultare ripugnante anche per loro, come per chiunque altro. In ogni caso, è importante che la NATO sia percepita come un attore responsabile e credibile in questo processo, così come lo è in altri ambiti, come il cyberspazio o lo spazio extra-atmosferico. In sintesi, quando si parla di IA, la NATO deve dare l'esempio" (Assemblea Parlamentare NATO, 2024, p. 13).

Riferimenti bibliografici

Abramson, J. (2017). *The Convention on Certain Conventional Weapons (CCW) At a Glance*. Arms Control Association. Disponibile a: https://www.armscontrol.org/factsheets/CCW.

ACHPR - African Commission on Human and Peoples' Rights. (2024). *Submission: Commissioner Solomon Ayele Dersso, Focal Point on the ACHPR Study on AI and Other Technologies on Lethal Autonomous Weapon Systems to the United Nations Secretary-General in terms of UN General Assembly Resolution 78/241*. Disponibile a: https://docs-library.unoda.org/General_Assembly_First_Committee_-Seventy-Ninth_session_(2024)/78-241-African_Commission-EN.pdf.

AFCEA Roma. (2023). *Conferenza: L'Intelligenza Artificiale in ambito militare: nuove soluzioni ed evoluzione d'impiego*. Casa dell'Aviatore, Roma. Disponibile a: https://www.afcearoma.it/eventi/anno-2023/138-l-intelligenza-artificiale-in-ambito-militare-nuove-soluzioni-ed-evoluzione-d-impiego.

Altman, J. & Quest, R. (2025). Europe "not in the AI race today", French President Macron says. *CNN*. Disponibile a: https://edition.cnn.com/2025/02/09/europe/france-macron-europe-ai-race-intl.

ANSA. (2024). *Crosetto, "Intelligenza artificiale priorità del governo".* Disponibile a: https://www.ansa.it/ansacom/notizie/economia/cybersec/2024/02/29/crosetto-intelligenza-artificiale-priorita-del-governo_d1f2244c-3539-4c3c-b32e-fefe4aa5f566.html.

Assemblea Parlamentare della NATO. (2024). *Special Report: NATO and Artificial Intelligence: navigating the challenges and opportunities*. Disponibile a: https://www.nato-pa.int/download-file?filename=/sites/default/files/20242/058%20STC%2024%20E%20rev.2%20fin%20-%20NATO%20AI%20-%20CLEMENT%20REPORT_0.pdf.

Automated Decision Research. (2025). *State positions*. Disponibile a: https://automatedresearch.org/state-positions/.

Blanchard, A., Boulanin, V., Bruun, L., & Goussac, N. (2025). *Dilemmas in the policy debate on autonomous weapon systems*. Stockholm International Peace Research Institute.

BPA - Ufficio stampa e informazione del Governo federale. (2025). *Il cancelliere Friedrich Merz incontra il CEO di NVIDIA Jensen Huang – La Germania rivendica la leadership nell'intelligenza artificiale* [tradotto dal tedesco]. Disponibile a: https://www.bundeskanzler.de/bk-de/aktuelles/bundeskanzler-friedrich-merz-trifft-

nvidia-ceo-jensen-huang-deutschland-mit-fuehrungsanspruch-bei-kuenstlicher-intelligenz-2354238.

CICR - Comitato Internazionale della Croce Rossa. (2021). *ICRC position on autonomous weapons systems.* International Committee of the Red Cross. Disponibile a: https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems.

CNN Transcripts. (2024). *Biden Addresses United Nations for Final Time as President; IDF Carries Out Second Wave of Strikes Against Hezbollah; Biden Speaks at U.N. as Middle East Conflict Intensifies*. Disponibile a: https://transcripts.cnn.com/show/ctw/date/2024-09-24/segment/02.

Consiglio dell'Unione Europea. (2021). *Conclusioni del Consiglio sulla sesta conferenza di revisione della convenzione sulla proibizione o la limitazione dell'uso di alcune armi convenzionali*. Disponibile a: https://data.consilium.europa.eu/doc/document/ST-13244-2021-INIT/it/pdf.

DFAE - Ufficio federale degli affari esteri. (2024) *Armi convenzionali*. Disponibile a: https://www.eda.admin.ch/eda/it/dfae/politica-estera/politica-sicurezza/armo-non-proliferazione/klassische-waffen.html.

DoS - US Department of State. (2023). *Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy*. Disponibile a: https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy-2/.

Ebo A. (2025). *Concluding remarks to the informal consultations on lethal autonomous weapons systems*. Discorso pronunciato in qualità di Direttore e Vice Alto Rappresentante per gli Affari del Disarmo presso l'UNODA. Disponibile a: https://front.un-arm.org/wp-content/uploads/2025/05/HR-remarks-Ade-12.05.2025.pdf.

Élysée. (2025). *Paris Declaration on Maintaining Human Control in AI enabled Weapon Systems*. Disponibile a: https://www.elysee.fr/emmanuel-macron/2025/02/11/paris-declaration-on-maintaining-human-control-in-ai-enabled-weapon-systems.

Gazzetta Ufficiale dell'Unione Europea. (2012). *Trattato sull'Unione Europea (versione consolidata)*. Disponibile a: https://eur-lex.europa.eu/resource.html?uri=cellar:2bf140bf-a3f8-4ab2-b506-fd71826e6da6.0017.02/DOC_1&format=PDF.

Gazzetta Ufficiale dell'Unione Europea. (2024). *Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale*. Disponibile a: https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=OJ:L_202401689.

GOVUK - Government of United Kingdom. (2025). *Prime Minister sets out blueprint to turbocharge AI*. _Press release. Disponibile a: https://www.gov.uk/government/news/prime-minister-sets-out-blueprint-to-turbocharge-ai.

Hoppenbrouwers, A. (2024). *The Global South and Autonomous Weapons Controls*. Arms Control Association. Disponibile a: https://www.armscontrol.org/act/2024-11/features/global-south-and-autonomous-weapons-controls.

Human Rights Watch. (2023). *Latin America and Caribbean Nations Rally Against Autonomous Weapons Systems*. Disponibile a: https://www.hrw.org/news/2023/03/06/latin-america-and-caribbean-nations-rally-against-autonomous-weapons-systems.

Ministero della Difesa. (2025). "La ricerca di nuove sicurezze: La difesa nazionale e la pace, fra incertezze UE ed egemonia" USA. Intervento del Ministro presso l'università di Padova. Disponibile a: https://www.difesa.it/primopiano/il-ministro-della-difesa-all-universita-di-padova-per-convegno-la-ricerca-di-nuove-sicurezze-la-difesa-nazionale-e-la-pace-fra-incertezze-ue-ed-egemonia-usa/73397.html.

NATO. (2021). *Summary of the NATO Artificial Intelligence Strategy*. Disponibile a: https://www.nato.int/cps/en/natohq/official_texts_187617.htm.

NATO. (2024). *Summary of NATO's revised Artificial Intelligence (AI) strategy*. Disponibile a: https://www.nato.int/cps/en/natohq/official_texts_227237.htm.

Perrin B., & Zamani M. (2025). *The Future of Warfare: National Positions on the Governance of Lethal Autonomous Weapons Systems*. Disponibile a: https://lieber.westpoint.edu/future-warfare-national-positions-governance-lethal-autonomous-weapons-systems/.

Perrin, B. (2025). *Lethal Autonomous Weapons Systems & International Law: Growing Momentum Towards a New International Treaty*. Disponibile a: https://www.asil.org/insights/volume/29/issue/1#:~:text=On%20December%202%2C%202024%2C%20the,Federation)%2C%20and%2015%20abstentions.

Pomfret, J. & Zhen, S. (2025). China's Xi calls for self-sufficiency in AI development amid U.S. rivalry. *Reuters*. Disponibile a: https://www.reuters.com/world/china/chinas-xi-calls-self-sufficiency-ai-development-amid-us-rivalry-2025-04-26/.

Principi guida adottati dal Gruppo di Esperti Governativi della CCW. (2019). *Report on emerging technologies in the area of lethal autonomous weapons systems*. Disponibile a: https://documents.unoda.org/wp-content/uploads/2020/09/CCW_GGE.1_2019_3_E.pdf.

Reaching Critical Will. (2025). *CCW Report: Civil society perspectives on the Group of Governmental Experts of the Convention on Certain Conventional Weapons on Lethal Autonomous Weapon Systems*. Disponibile a: https://reachingcriticalwill.org/disarmament-fora/ccw/2025/laws/ccwreport/17326-ccw-report-vol-13-no-2.

Rete Italiana Pace Disarmo. (2024). *Armi autonome: il voto ONU stimoli negoziati per un trattato*. Disponibile a: https://retepacedisarmo.org/stop-killer-robots/2024/12/armi-autonome-il-voto-onu-stimoli-negoziati-per-un-trattato/.

Roll Call. (2025). Remarks: Donald Trump Announces AI Infrastructure Initiative. Disponibile a: https://rollcall.com/factbase/trump/transcript/donald-trump-remarks-infrastructure-investment-ai-january-21-2025/.

Santagata, E. & Melegari, A. (2017). Putin esorta gli studenti a dedicarsi all'Intelligenza Artificiale. *Analisi Difesa*. Disponibile all'indirizzo: https://www.analisidifesa.it/2017/09/putin-esorta-agli-studenti-a-dedicarsi-allintelligenza-artificiale/.

SKR - Campagna Internazionale Stop Killer Robots. (2025). *Our member organisations*. Disponibile a: https://www.stopkillerrobots.org/a-global-push/member-organisations/.

Spazian A., Holland Michel A., & Anand A. (2021). *UNIDIR on Lethal Autonomous Weapons*. Disponibile a: https://unidir.org/wp-content/uploads/2023/05/UNIDIR-on-Lethal-Autonomous-Weapons-Final.pdf.

Sterling, B. (2020). Accelerate the development of military intelligentization. *Wired*. Disponibile a: https://www.wired.com/beyond-the-beyond/2020/01/accelerate-development-military-intelligentization-/.

Taddeo M., McNeish D., Blanchard A., & Edgar E. (2021). *Ethical principles for artificial intelligence in national defence*. *Philosophy & Technology, 34*:1707–1729.

The White House. (2025a). *Winning the Race. America's AI Action Plan*. Disponibile a: https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf.

The White House. (2025b). *Public Comment Invited on Artificial Intelligence Action Plan*. Disponibile a: https://www.whitehouse.gov/briefings-statements/2025/02/public-comment-invited-on-artificial-intelligence-action-plan/.

UNGA - United Nations General Assembly. (2023). *Resolution 78/241 on lethal autonomous weapons systems.* Disponibile a: https://docs.un.org/en/A/RES/78/241.

UNGA - United Nations General Assembly. (2024a). *Resolution 79/62 on lethal autonomous weapons systems.* Disponibile a: https://docs.un.org/A/RES/79/62. Risultati votazione. Disponibile a: https://reachingcriticalwill.org/images/documents/Disarmament-fora/1com/1com24/votes-ga/408DRXLIII_.pdf.

UNGA - United Nations General Assembly. (2024b). *Lethal autonomous weapons systems Report of the Secretary-General.* Disponibile a: https://docs-library.unoda.org/General_Assembly_First_Committee_-Seventy-Ninth_session_(2024)/A-79-88-LAWS.pdf.

UNODA - United Nations Office of Disarmament Affairs. (2023). *Lethal Autonomous Weapon Systems (LAWS)*. Disponibile a: https://disarmament.unoda.org/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/.

UNSG - United Nations Secretary General. (luglio 2023). *Our Common Agenda: Policy Brief 9 – A New Agenda for Peace*. Disponibile a: https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-new-agenda-for-peace-en.pdf.

Van den Boogaard, J. (2024). *Warning! Obstacles Ahead! The Regulation of Autonomous Weapons Systems in the GGE LAWS*. Disponibile a: https://opiniojuris.org/2024/03/04/warning-obstacles-ahead-the-regulation-of-autonomous-weapons-systems-in-the-gge-laws/.

Parte IV – La sfida sociale

Capitolo 6 – L'impegno della società civile in Italia e all'estero: chiese, associazioni, movimenti (2020-2025)

6.1. Premessa

Abbiamo visto come, negli ultimi anni, il processo di autonomizzazione della guerra abbia sollevato profonde preoccupazioni etiche, legali e sociali. In particolare, il nuovo fenomeno ha alimentato un acceso dibattito internazionale e stimolato una vasta mobilitazione della società civile per prevenire lo sviluppo e la diffusione incontrollati della IA militare. In questo scenario, la società civile ha assunto un ruolo crescente nel promuovere consapevolezza, riflessione critica e richiesta di responsabilità, contribuendo a mantenere viva l'attenzione pubblica e istituzionale sulle implicazioni umane e normative dell'uso militare delle tecnologie autonome.

6.2. Advocacy e mobilitazione politica

Nel periodo 2020-2025, l'impegno della società civile si è articolato in una vasta gamma di iniziative pubbliche, accademiche, istituzionali e anche artistiche. Il quinquennio in esame ha visto l'intensificazione delle attività di *advocacy* diretta a decisori politici e organismi internazionali, con l'obiettivo di influenzare il dibattito normativo e promuovere l'adozione di un trattato vincolante sulle armi autonome.

6.2.1. La Campagna internazionale Stop Killer Robots (SKR)

Pur nella pluralità delle sue articolazioni e relative differenze, la società civile condivide un obiettivo comune e imprescindibile nella lotta contro l'impiego dei *Sistemi d'Arma Automi* (*AWS*): assicurare che l'automazione delle decisioni militari non comprometta i diritti umani fondamentali, la dignità delle persone e la stabilità globale. Sebbene ciascuna parte operi con approcci e strumenti diversificati, tutte queste realtà convergono in un appello urgente per una regolamentazione internazionale capace di frenare l'uso incontrollato di tecnologie belliche sempre più autonome.

L'epicentro del movimento proibizionista è rappresentato dalla Campagna internazionale *Stop Killer Robots* (*SKR*), una coalizione di 270 organizzazioni della società civile attive in più di 70 Paesi (SKR, 2025a), di cui diverse con sede in Paesi membri della NATO e/o dell'Unione Europea (v. appendice 1). Lanciata nell'aprile del 2013, la Campagna sostiene che il diritto esistente non è più sufficiente a tutelare l'umanità dai rischi etici e legali che le armi autonome comportano. Le sue argomentazioni principali ruotano attorno alla necessità di preservare il controllo umano nel ricorso alla forza e di rispettare il *Diritto Internazionale Umanitario* (*DIU*) nelle azioni belliche.

Tra il 2018 e il 2021, la Campagna *SKR* ha vissuto una fase di crescita. Nel 2022-2023, dopo essersi rafforzata nelle reti territoriali (nazionali), ha contribuito sulla scena internazionale all'adozione di una prima risoluzione dell'Assemblea Generale ONU. Nel 2024, ha partecipato al consolidamento dell'impegno alla regolamentazione degli *AWS,* grazie a una seconda risoluzione ONU che ha avviato un processo di consultazioni informali espressamente aperto alla società civile. In questo contesto, il crescente sostegno di molti Stati a un trattato in materia e il riconoscimento dell'*expertise* della società civile hanno confermato l'efficacia dell'approccio della Campagna. Nel 2025, le azioni si concentrano sul rafforzamento dell'*advocacy* parlamentare nei singoli Paesi con l'obiettivo di consolidare il consenso politico necessario ad avviare veri e propri negoziati internazionali. A supporto di queste attività, la Campagna *SKR* mette a disposizione fondi "simbolici" per i propri membri (fino a 5.000 dollari per singolo progetto), destinati ad azioni di sensibilizzazione e pressione politica al fine di arrivare a uno strumento giuridicamente vincolante sulla IA militare.

Tra le attività di *advocacy* verso le istituzioni da parte della Campagna, è da ricordare la pubblicazione di *Stopping Killer Robots: A Guide for Policy Makers* (SKR, 2021a), una guida in vista dello sviluppo di un nuovo trattato internazionale contro i pericoli posti dalla IA militare. A questa guida ha fatto seguito, il 7 luglio 2022, il lancio di un *Parliamentary Pledge*. L'iniziativa invita i parlamentari di tutti i Paesi rappresentati nelle Nazioni Unite a dichiarare il proprio impegno contro l'automazione letale degli armamenti, a sostenere il mantenimento di un controllo umano nei processi decisionali e a promuovere l'avvio di negoziati per uno strumento giuridicamente vincolante sugli *AWS* (SKR, 2024a). Attualmente, i firmatari sono 132, con una forte partecipazione da parte dei parlamentari austriaci (47), seguiti da francesi (20) e tedeschi (17); tre gli italiani (SKR, 2025b). In aggiunta, nel 2024 la Campagna *SKR* ha rilasciato una guida per i propri membri, fornendo loro strumenti di *advocacy* con le istituzioni (SKR, 2024b). Queste iniziative, coordinate centralmente ma implementate a livello nazionale, hanno favorito una crescente mobilitazione politica nei singoli Paesi, contribuendo a rafforzare le attività di *advocacy* delle associazioni aderenti.

6.2.2. Organizzazioni internazionali e associazioni per il controllo delle armi autonome

È evidente che una delega delle decisioni di vita o di morte a sistemi privi di discernimento umano rappresenta una potenziale minaccia ai principi fondamentali del *DIU*. In questo dibattito, un ruolo centrale è esercitato dal Comitato Internazionale della Croce Rossa (CICR), il quale insiste sulla necessità che gli Stati debbano non solo evitare violazioni del *DIU,* come la distinzione tra civili e combattenti o il principio di proporzionalità negli attacchi, ma anche garantire che l'uso della forza resti sotto *Controllo Umano Significativo* (*CUS*). Si tratta di un vero e proprio "obbligo positivo": non è sufficiente vietare l'abuso, ma occorre che un sistema sia progettato, testato e

impiegato in modo tale da permettere che l'intervento umano sia consapevole e responsabile. Questo principio implica che gli Stati devono:

- Progettare sistemi che consentano la supervisione e l'intervento umano;
- Prevedere procedure istituzionali per l'approvazione e il monitoraggio dei sistemi autonomi;
- Garantire una catena di responsabilità chiara lungo tutto il ciclo operativo dell'arma (England, 2016, pp. 4-5).

In qualità di garante del *DIU*, il CICR avanza proposte di nuove normative con la massima cautela. Allo stesso tempo, esso è fermamente impegnato a favorire l'evoluzione del diritto affinché le norme attuali non vengano compromesse. L'obiettivo del Comitato Internazionale della Croce Rossa è garantire che le protezioni per le persone coinvolte nei conflitti siano pienamente rispettate e, laddove necessario, rafforzate per far fronte alle innovazioni incorporate nelle armi e nelle conseguenti modalità di combattimento (CICR, 2022). È inoltre da ricordare il decisivo apporto della Croce Rossa alla definizione dell'approccio a due livelli nella regolamentazione delle *AWS*[1].

Accanto al ruolo di una vera e propria istituzione come Croce Rossa, non è da trascurare il diffuso impegno di una serie di ONG e associazioni attive a livello internazionale.

I primi segnali di allarme riguardanti l'impiego degli *AWS* risalgono al 2012, quando Human Rights Watch e la International Human Rights Clinic della Harvard Law School hanno pubblicato *Losing Humanity: The Case Against Killer Robots*. Il Rapporto analizza in modo critico l'uso degli *AWS* e rappresenta uno dei primi appelli strutturati contro l'uso di queste tecnologie. Il documento ha determinato l'inizio delle attività della Campagna *Stop Killer Robots* (*SKR*) ed è stato fondamentale per far emergere il tema, rappresentando (ancora oggi) uno dei punti di riferimento del dibattito etico e giuridico sulla questione (Human Rights Watch & International Human Rights Clinic, 2012).

Amnesty International ha rivolto particolare attenzione all'impiego delle armi autonome nell'ambito dei compiti di polizia e nell'ambito della sicurezza interna. L'organizzazione mette in guardia contro il rischio che queste tecnologie, soprattutto se abbinate a sistemi di riconoscimento facciale o algoritmi discriminatori, possano violare seriamente i diritti delle persone, colpendo in modo sproporzionato categorie già vulnerabili (Amnesty International, 2015; Amnesty International Italia, 2021).

Altre realtà della società civile, come *Africa Teen Geeks*, si sono espresse con forza a favore di una regolamentazione più stringente. L'organizzazione, attiva nel campo dell'educazione tecnologica in Africa, evidenzia come la corsa agli armamenti basata sulla IA possa destabilizzare equilibri geopolitici già fragili. Per questo promuove un dialogo multilaterale che coinvolga industria, scienza, governi e cittadini, finalizzato a definire regole chiare che preservino i diritti umani e la sicurezza collettiva (UNSG, 2024).

[1] V. *supra*. par. 5.5.3.

Infine, l'*Arms Control Association* porta l'attenzione su uno scenario particolarmente inquietante: l'autonomizzazione dei sistemi di Comando e Controllo delle armi nucleari. L'associazione sottolinea il pericolo che decisioni algoritmiche in contesti ad altissima tensione strategica possano sfociare in escalation catastrofiche, anche in assenza di intenzionalità. Per evitare rischi inaccettabili per la sicurezza globale, propone la creazione di un organismo internazionale di monitoraggio sull'uso della IA nei sistemi militari, in particolare quelli nucleari, ribadendo che ogni decisione relativa all'uso di *Armi di Distruzione di Massa* (*AMD*) deve restare saldamente in mano umana (Rautenbach, 2022).

Al coro di voci critiche si aggiungono quelle di singoli scienziati e imprenditori attivi nella robotica e nella IA, tra cui quella di Elon Musk. Già nel 2015 il discusso imprenditore, protagonista un decennio più tardi di una controversa collaborazione con il presidente degli Stati Uniti Donald Trump, ha sottoscritto una lettera aperta[2] in cui si sottolinea il rischio di innescare una "terza rivoluzione nella guerra" (Future of Life Institute, 2017).

La guerra automatizzata, o in futuro addirittura *autonomizzata*, sebbene tecnicamente avanzata, potenzialmente non è meno devastante di quella tradizionale. Le armi autonome e le tecnologie come la sorveglianza totale, gli attacchi informatici e, sul campo, i sistemi d'arma guidati dagli algoritmi stanno cambiando la natura dei conflitti, ma le vittime rischiano di rimanere sempre esseri umani.

Il dibattito sugli *AWS* rappresenta uno dei nodi più complessi e urgenti del *DIU* contemporaneo. La tecnologia evolve a un ritmo che sfida le strutture normative esistenti, mentre le implicazioni etiche, giuridiche e umanitarie restano profondamente controverse. Un'azione collettiva e concertata a livello internazionale sarà essenziale per garantire che la IA non comprometta i valori fondamentali di dignità, responsabilità e sicurezza umana a livello globale.

6.2.3. Interlocuzione con le istituzioni e incidenza politico-diplomatica

Proprio grazie all'azione concertata della società civile, a cominciare dalle azioni promosse dalla Campagna *SKR*, diverse associazioni e organizzazioni non governative hanno focalizzato la propria attività di *advocacy* sul dialogo con le istituzioni a livello sia nazionale (parlamenti) sia internazionale (ONU), al fine di aumentare il supporto e accelerare il processo verso un trattato internazionale.

Il *Future of Life Institute* (con sedi in Belgio e negli Stati Uniti e finanziato da Elon Musk) sottolinea i dilemmi morali e giuridici legati alla delega di decisioni letali a entità artificiali. L'istituto ha diffuso, il 5 agosto 2024 e in una versione aggiornata il 17 settembre 2024, il testo intitolato *A Diplomat's Guide to Autonomous Weapons Systems*, una guida sintetica pensata per i diplomatici coinvolti nei negoziati sul disarmo. La pubblicazione illustra le principali problematiche legate alla IA militare, le posizioni degli

[2] La lettera è stata firmata da oltre cento specialisti provenienti da più di 20 Paesi.

Stati, i rischi percepiti, e le tappe chiave del percorso verso un trattato, contribuendo a una maggiore consapevolezza all'interno delle sedi istituzionali (Future of Life Institute, 2024a).

Sempre su scala globale, una dichiarazione congiunta a favore del disarmo umanitario è stata approvata nell'ottobre 2024 da 90 organizzazioni della società civile[3] e letta pubblicamente da Erin Hunt (*Mines Action Canada*) durante la sessione della Prima Commissione sul Disarmo dell'Assemblea Generale delle Nazioni Unite. L'intervento ha rappresentato un momento significativo nell'impegno per aumentare la visibilità e la pressione collettiva della società civile nelle sedi multilaterali (Rete Italiana Pace e Disarmo, 2024).

Per uno sguardo sui singoli Paesi, nel biennio 2021-2022, la *United Nations Association* (*UNA-UK*), in qualità di coordinatrice del ramo inglese della Campagna *SKR*, ha intensificato il lavoro con parlamentari di tutti i principali partiti, contribuendo a oltre 20 interrogazioni scritte, 10 interventi orali in aula e tre eventi pubblici dedicati al tema. Tra le iniziative promosse: un evento sul legame tra investimenti "opachi" e robot killer, la diffusione di analisi tecniche utilizzate nei dibattiti alla Camera dei Lords, e il sostegno a interrogazioni parlamentari sulla strategia del Ministero della Difesa in materia di IA (UNA-UK, 2022; UK SKR, 2022).

Un'altra organizzazione britannica, la *Article 36*, pone l'accento sul rischio che l'inserimento della IA nei sistemi d'arma comprometta la capacità di prevederne gli effetti, specialmente quando tali sistemi operano su dati incompleti o distorti. Nel luglio 2024, l'organizzazione ha pubblicato un comunicato rivolto al nuovo Governo britannico, sottolineando la necessità di sostenere un processo internazionale finalizzato alla proibizione della IA militare e fornendo raccomandazioni concrete per rafforzare il coinvolgimento del Regno Unito nei negoziati multilaterali (Article 36, 2024a).

In Italia, l'8 ottobre 2024, in un'ottica di sensibilizzazione politica dei vertici istituzionali, una delegazione dell'Istituto di Ricerche Internazionali Archivio Disarmo - IRIAD, in collaborazione con la Rete Italiana Pace e Disarmo, è stata ricevuta dal Presidente della Repubblica Sergio Mattarella per esporre i rischi etici e strategici connessi allo sviluppo delle armi autonome. All'incontro era presente il professor Peter Asaro, vicepresidente dell'*International Committee for Robot Arms Control* (*ICRAC*) ed esponente della Campagna *SKR*, in un'ottica di collegamento tra attivismo internazionale e istituzioni nazionali (9Colonne, 2024). Successivamente, il 29 ottobre, Archivio Disarmo, Rete Italiana Pace e Disarmo e Amnesty International Italia sono intervenuti in un'audizione informale davanti alla Commissione Affari Esteri della Camera dei Deputati per presentare la Campagna volta a promuovere, in sede ONU, limitazioni all'uso degli *AWS* (Camera dei Deputati, 2024).

Anche le comunità religiose hanno assunto un ruolo sempre più visibile nel dibattito sulle armi autonome. Il 27 agosto 2024, in concomitanza con la sessione del *Gruppo di Esperti Governativi* (*GEG*) dell'ONU sugli *AWS* (tenutasi dal 26 al 30 agosto a Ginevra),

[3] Per l'elenco dei firmatari, si rimanda a Rete Italiana Pace e Disarmo, 2024.

il Consiglio Ecumenico delle Chiese (CEC) ha promosso, con il sostegno della Campagna *SKR*, un evento interreligioso dal titolo *Faith communities take a stand against lethal autonomous weapons*. Rappresentanti di Cristianesimo, Islam, Soka Gakkai e fede Bahá'í hanno sottolineato che "il valore della vita umana è inestimabile e le decisioni di vita o di morte non possono essere delegate alle macchine" (CEC, 2024a), riaffermando così la sacralità della vita come principio comune alle diverse tradizioni religiose.

Questa posizione è stata ulteriormente consolidata a novembre 2024 quando il CEC ha pubblicato un documento ufficiale che, oltre a ribadire la sua opposizione categorica alla IA militare, ha esortato governi e comunità internazionale a "investire molto di più, da parte dei governi europei e dell'intera comunità internazionale, nella ricerca e nella promozione della pace e nel rafforzamento dei processi di risoluzione nonviolenta dei conflitti e di riconciliazione, piuttosto che nell'inasprimento degli scontri e delle divisioni" (CEC, 2024b). Inoltre, già nel 2021 il CEC ha pubblicato una guida (*Killer Robots: A Campaign Guide for Churches*) per sensibilizzare le comunità cristiane sui rischi legati ai robot killer, incoraggiando un impegno attivo per un divieto preventivo del loro sviluppo (CEC, 2022). La guida, disponibile in arabo, francese, inglese, portoghese, spagnolo e tedesco, era stata presentata durante un webcast coordinato da Emily Welty (vice-moderatrice della Commissione delle Chiese per gli affari internazionali del Consiglio mondiale delle Chiese), con la partecipazione di Isabelle Jones (responsabile della campagna di sensibilizzazione per *SKR*), il reverendo Kolade Fadahunsi (direttore dell'Istituto di Chiesa e Società del Consiglio cristiano della Nigeria), il vescovo Michael Vorster (direttore degli Affari ecumenici della Chiesa metodista dell'Africa meridionale) e il vescovo Christopher Cocksworth (vescovo di Coventry).

Da parte sua, in diverse occasioni la Chiesa cattolica è intervenuta con fermezza. Papa Francesco, durante la sua partecipazione al G7 il 14 giugno 2024, ha espresso la sua preoccupazione per il predominio del "paradigma tecnocratico" e sollecitato l'adozione di una "sana politica" per promuovere il "buon uso" degli algoritmi. Inoltre, ha ribadito che "nessuna macchina dovrebbe mai scegliere se togliere la vita a un essere umano" (Vatican News, 2024). Successivamente, in un messaggio all'incontro multireligioso *AI Ethics for Peace* svoltosi a Hiroshima il 9-10 luglio 2024, il Pontefice ha chiesto "un fattivo impegno per tutelare la dignità umana in questa nuova stagione di uso delle macchine" e ha rilanciato la richiesta di bandire l'uso delle armi autonome (Campisi, 2024).

Questa preoccupazione era stata riconosciuta formalmente nel 2020 con la pubblicazione della *Rome Call for AI Ethics*, promossa dalla Pontificia Accademia per la Vita. Il documento, incentrato sulla tutela della dignità umana, sottolineava la necessità di mantenere sempre il *CUS* sui sistemi decisionali, una posizione rilevante anche nel dibattito sulle armi autonome. Tra i firmatari figurano attori istituzionali ed economici di primo piano come Microsoft, IBM e FAO, a testimonianza dell'ampio riconoscimento internazionale del documento. Per sostenere e diffondere questi principi, nel 2021 è stata istituita anche la *Fondazione RenAIssance*, che continua a promuovere la *Rome Call* come

quadro etico di riferimento per lo sviluppo della IA, inclusi gli ambiti militari[4] (Rancilio, 2025).

6.3. Conoscenza e consapevolezza: informare per incidere

Per poter influire sui decisori, l'*advocacy* in materia di armi autonome deve fare ricorso a solide basi di conoscenza. A sua volta la conoscenza non è solo un fine, tra i più elevati propri della civiltà umana, ma anche uno strumento politico e culturale. La produzione di contenuti analitici e critici si intreccia con la loro diffusione pubblica, alimentando un ciclo virtuoso tra ricerca, divulgazione e mobilitazione. È in questa interazione tra approfondimento e partecipazione che si radica la capacità della società civile di incidere in modo duraturo.

6.3.1. Produzione di conoscenza e ricerca

La produzione di conoscenza è fondamentale per alimentare un dibattito informato e approfondito sulla IA militare. Nel quinquennio considerato, la società civile ha investito con continuità nella produzione di contenuti analitici, informativi e divulgativi, contribuendo a consolidare il dibattito sui rischi associati all'uso e alla proliferazione degli *AWS*. Attraverso report tecnici, articoli di approfondimento, materiali audiovisivi e format educativi, organizzazioni e centri di ricerca hanno alimentato la riflessione critica sul tema e dato sostegno alle richieste di regolamentazione.

Sul piano internazionale, nel 2021 la Campagna *SKR* ha creato un gruppo di monitoraggio e ricerca che ha poi lanciato nel 2022 con il nome di *Automated Decision Research*. Il team produce rapporti, briefing e schede informative[5].

Significativi contributi sono provenuti da specifiche prospettive metodologiche e sociali, quali ad esempio il femminismo. Nel 2020, *Reaching Critical Will* ha pubblicato, con il supporto della Campagna *SKR*, una serie di testi analitici intitolata *Feminist perspectives on autonomous weapon systems* (Reaching Critical Will, 2020). In particolare, i paper *Autonomous weapons and patriarchy* e *Autonomous weapons and gender-based violence*, redatti da Ray Acheson, propongono una lettura critica delle armi autonome quale espressione di strutture di potere patriarcali e militarizzate. I testi esaminano le implicazioni in termini di violazioni dei diritti umani e delle libertà fondamentali, evidenziando come l'astrazione della violenza veicolata da tali tecnologie

[4] Sebbene al momento della stesura di questo capitolo (maggio 2025) non siano stati ancora rilasciati riferimenti espliciti sulle armi autonome, papa Leone XIV ha indicato la volontà di proseguire sulla linea del predecessore Francesco sul tema della IA. Il nuovo papa ha spiegato la scelta del proprio nome sostenendo che "Papa Leone XIII, con la storica Enciclica *Rerum Novarum* affrontò la questione sociale nel contesto della prima grande Rivoluzione industriale. Oggi la Chiesa offre a tutti il suo patrimonio di dottrina sociale per rispondere a un'altra rivoluzione industriale e agli sviluppi dell'Intelligenza Artificiale, che comportano nuove sfide per la difesa della dignità umana, della giustizia e del lavoro" (Bonanata, 2025).

[5] Per le singole pubblicazioni si rimanda a SKR, 2021b.

contribuisca a normalizzare la disumanizzazione e a perpetuare una cultura dell'impunità, in particolare nei confronti delle forme di violenza di genere. Questa prospettiva teorica amplia il quadro interpretativo, suggerendo che il contrasto agli *AWS* vada affrontato non solo sul piano tecnico-normativo, ma anche come lotta contro le logiche sistemiche di violenza e di discriminazione.

Nel marzo 2024 il Comitato Internazionale della Croce Rossa (CICR) e la *Geneva Academy of International Humanitarian Law and Human Rights* (Svizzera) hanno pubblicato l'*Expert Consultation Report on AI and Related Technologies in Military Decision-Making on the Use of Force in Armed Conflicts*. Il rapporto, frutto del lavoro di Anna Greipl (*Geneva Academy*) con Neil Davison e Georgia Hinds (CICR), analizza lo stato attuale della IA militare e le relative conseguenze per il *DIU* (Geneva Academy of International Humanitarian Law and Human Rights & CICR, 2024). Alcuni mesi dopo, ad agosto, la *Geneva Academy of International Humanitarian Law and Human Rights* (2024) ha pubblicato anche un *research brief* che evidenzia i rischi posti dalla proliferazione degli *AWS* per gli attori statali e non statali, concentrandosi sulla sicurezza internazionale e sulle preoccupazioni legali.

Nel Regno Unito, l'8 febbraio 2024 il *Royal United Services Institute* (*RUSI*) ha pubblicato il paper *Assessing Autonomous Weapons as a Proliferation Risk*, presentato poi durante l'evento *The Proliferation Risk of Lethal Autonomous Weapons - Paper Launch*. L'incontro ha coinvolto esperti di difesa e sicurezza per discutere le implicazioni strategiche legate alla diffusione incontrollata di queste tecnologie (RUSI, 2024). Sempre nel Regno Unito, nel dicembre 2024, *Article 36* ha diffuso l'analisi *Opportunities after the UNGA Resolution on Autonomous Weapons: Moving Toward a New Treaty*, incentrata sulle prospettive di un nuovo trattato internazionale alla luce della risoluzione adottata dall'Assemblea Generale delle Nazioni Unite (Article 36, 2024b).

Negli Stati Uniti, nel novembre 2024 l'organizzazione *Public Citizen* ha pubblicato un rapporto di denuncia sull'industria dei robot killer, segnalando una "sconsiderata e pericolosa corsa agli armamenti" condotta negli USA in relazione agli *AWS* e richiedendo una pressione più forte dell'opinione pubblica sul Congresso e sul Governo federale (Public Citizen, 2024).

In ambito accademico e giuridico, Human Rights Watch e la Clinica internazionale dei diritti umani della Harvard Law School hanno pubblicato il rapporto *A Hazard to Human Rights: Autonomous Weapons Systems and Digital Decision-Making*, in vista della prima riunione dell'Assemblea Generale dell'ONU sugli *AWS*, tenutasi a New York il 12-13 maggio 2025. Il rapporto (61 pagine) sostiene che l'impiego di armi autonome, basate su sensori anziché input umani, minaccia i diritti alla vita, alla privacy, alla riunione pacifica, al risarcimento, nonché i principi della dignità umana e della non discriminazione (Docherty, 2025).

L'impegno contro le armi autonome sta avendo riscontro anche in Spagna, dove il *Centro de Educación e Investigación para la Paz* ha dato spazio ai rischi degli *AWS* nel suo report annuale 2019-2020 *Riesgos globales y multilateralismo: el impacto de la COVID-19* (Centro de Educación e Investigación para la Paz, 2020) e 2023-2024

Oportunidades de paz y lógicas de guerra (Centro de Educación e Investigación para la Paz, 2024).

In Belgio, il *Groupe de recherche et d'information sur la paix et la sécurité* (*GRIP*) che ha dedicato molta attenzione ai "*robots tueurs*" pubblicando il paper *Encadrement des robots tueurs: vers des négociations en dehors de la Convention sur certaines armes classiques?* (Bannenberg & GRIP, 2022).

Analoghe attività di studio hanno avuto luogo anche in Italia, animando il dibattito pubblico nel periodo 2020-2025. Due report in particolare hanno fornito una base di analisi e riflessione nell'ambito nazionale ed europeo. Il primo è *La questione delle armi letali autonome e le possibili azioni italiane ed europee per un accordo internazionale*, rapporto di ricerca presentato nel 2020 da Archivio Disarmo, in collaborazione con l'Unione degli Scienziati per il Disarmo (*USPID*), il Ministero degli Affari Esteri e della Cooperazione Internazionale (IRIAD, 2020). Il Rapporto fa il punto sullo sviluppo e sull'utilizzo delle armi autonome e sulle possibili azioni italiane ed europee per un accordo internazionale in materia, fornendo una panoramica delle implicazioni legali, di sicurezza ed etiche legate allo sviluppo e all'utilizzo di questi sistemi d'arma.

Il secondo contributo da ricordare è il rapporto *Man in the Loop. Ricerca e sviluppo dei sistemi d'arma autonomi in Italia*, pubblicato nel 2023 dalla piattaforma informativa info.nodes (Carrer *et al.*, 2023). Il documento analizza le dinamiche di ricerca, di sviluppo e di impiego di tecnologie militari autonome nel contesto italiano, offrendo una mappatura aggiornata delle collaborazioni tra l'industria della difesa, il mondo accademico e le istituzioni pubbliche. Il lavoro si distingue per un approccio critico e documentato, in linea con la prospettiva promossa dalla Campagna *SKR*, di cui info.nodes è partner.

6.3.2. Divulgazione di conoscenza e attivismo culturale

Nel periodo 2020-2025, una parte significativa dell'impegno della società civile si è concretizzata in momenti di confronto, dialogo e scambio tra esperti, attivisti e pubblico, volti ad approfondire le implicazioni strategiche, etiche e sociali delle armi autonome. Questi appuntamenti, promossi in ambiti accademici, istituzionali e informali, hanno favorito il consolidamento di reti e la sensibilizzazione di una pluralità di attori.

6.3.2.1. Iniziative accademiche e seminari tematici

In ambito accademico e scientifico, numerose università europee hanno ospitato momenti di studio e confronto multidisciplinare, offrendo contributi rilevanti alla riflessione pubblica sull'impiego della IA in ambito militare.

Il 31 gennaio 2024, l'Università di Oxford (Regno Unito) ha organizzato il seminario *AI and the Military: Beyond Autonomous Weapons to Strategic Enterprise Transformation*, incentrato sull'utilizzo dell'Intelligenza Artificiale nelle operazioni

militari e sull'importanza dell'allineamento dei valori tra essere umano e macchina in ambito strategico (Università di Oxford, 2024).

In Italia il Dipartimento di Scienze Sociali ed Economiche e il Dipartimento di Fisica della Sapienza Università di Roma, in collaborazione con Archivio Disarmo, il 6 dicembre 2023 hanno organizzato il convegno *Intelligenza Artificiale: pace o guerra?*. Alla tavola rotonda è intervenuto Giorgio Parisi, Premio Nobel per la fisica nel 2021, sottolineando "negli ultimi 20 anni sempre più scarse sono state le azioni volte al perseguimento della pace […]. Esistono problematiche per le quali bisogna studiare e prepararsi. L'Intelligenza Artificiale è una scoperta che cambia le cose, che ha maggior potere sull'uomo, ma il compito della società è controllare questo potere" (IRIAD, 2023). Successivamente, l'11 ottobre 2024 presso la Facoltà di Scienze Politiche dello stesso ateneo, Archivio Disarmo ha promosso il convegno "*Intelligenza" delle macchine e follia della guerra* (IRIAD, 2024a)

Sempre in Italia, il 20 maggio 2024, presso il Dipartimento di Fisica dell'Università di Pisa, si è svolto il seminario *La militarizzazione dell'Intelligenza Artificiale e i sistemi di arma autonomi*, condotto da Guglielmo Tamburrini dell'Università di Napoli Federico II e dedicato alle implicazioni etiche e ai rischi legati all'automazione bellica (Dipartimento di Fisica - Università di Pisa, 2024).

6.3.2.2. Spazi pubblici di confronto promossi dalla società civile

Anche in ambito extra-accademico, la società civile ha promosso occasioni di discussione pubblica, volte a stimolare il dibattito e a sostenere la necessità di una regolamentazione efficace della IA militare.

Il 25 giugno 2021, in Belgio, il *GRIP* e Pax Christi Vlaanderen hanno organizzato un webinar bilingue intitolato *Robots tueurs: la Belgique plaidera-t-elle pour une interdiction?*, che ha raccolto rappresentanti istituzionali, accademici e della società civile per discutere i rischi associati allo sviluppo tecnologico in ambito militare e le prospettive diplomatiche del Belgio in materia di disarmo (GRIP, 2021). Inoltre, il 15 marzo 2022, il *GRIP* ha promosso un ulteriore webinar, *L'interdiction des robots tueurs: que peut faire la Belgique?*, volto a discutere le posizioni ufficiali del Governo belga e le raccomandazioni espresse dalla Campagna *SKR* nel contesto nazionale (GRIP, 2022).

In Austria, il 16 aprile 2024 il *Vienna Center for Disarmament and Non-Proliferation* ha ospitato il webinar *Lethal Autonomous Weapon Systems: Where Are We and What's Next?*, durante il quale è stata ribadita la necessità urgente di un trattato internazionale vincolante per regolare l'uso della IA militare (Vienna Center for Disarmament and Non-Proliferation, 2024). Inoltre, il 28 aprile 2024, a Vienna, la Campagna *SKR*, la Croce Rossa Austriaca e l'associazione PAX (già Pax Christi) hanno promosso il forum della società civile *Action at the Crossroads: Autonomous Weapons Systems and the Challenge of Regulation*, organizzato in concomitanza con la conferenza intergovernativa sugli *AWS* promossa dal Ministero austriaco per gli Affari Europei e Internazionali (2024). L'evento

ha visto la partecipazione di attivisti, esperti e rappresentanti di associazioni religiose, segnando un momento di forte mobilitazione internazionale.

In Svezia, a novembre 2024, il programma *Wallenberg AI, Autonomous Systems and Software Program – Humanity and Society* (*WASP-HS*) ha promosso, all'interno della conferenza *AI for Humanity and Society 2024*, il workshop *In Defense of Dignity in the Face of the Lethal Use of Artificial Intelligence*. Il dibattito si è concentrato sulle dinamiche di deumanizzazione della guerra, proponendo prospettive etiche e tecnofemministe nel governo delle tecnologie emergenti (WASP-HS, 2024).

Tra le attività più recenti a livello internazionale, nell'aprile 2025, il programma *Reaching Critical Will* della *Lega Internazionale delle Donne per la Pace e la Libertà* (*WILPF*) ha organizzato un incontro virtuale per i suoi membri. L'evento, dal titolo *WeAreWILPF: Autonomous Weapons and Engaging in WILPF's Work on Disarmament*, ha fornito aggiornamenti sui lavori del *GEG* dell'ONU, sottolineando la necessità di una prospettiva femminista e intergenerazionale nel disarmo, con il contributo del gruppo Young *WILPF* e del neocostituito *Disarmament Working Group* (WILPF, 2025).

Anche in Italia si sono svolti importanti confronti culturali in ambio extra-accademico. Nel novembre 2022, a Torino, il Centro Studi Sereno Regis ha organizzato *L'ABC della PACE: scienza e tecnologia*, in collaborazione con Archivio Disarmo e Rete Italiana Pace e Disarmo. L'appuntamento ha analizzato i rischi connessi alla militarizzazione della tecnologia e promosso una riflessione sullo sviluppo scientifico orientato alla pace (Rete Italiana Pace e Disarmo, 2025a).

6.3.2.3. Coinvolgimento dei giovani e strategie di divulgazione informale

Numerose iniziative sono state inoltre rivolte al coinvolgimento di giovani, attivisti e cittadini non specialisti, nell'ottica di rendere il tema della IA militare accessibile a un pubblico più ampio.

Tra queste, il 12 dicembre 2020 si è svolta online la *Global Youth Conference on Fully Autonomous Weapons*, primo evento globale giovanile sul tema, organizzato dall'*International Student Conference* (*ISC*) e dalla Campagna *SKR*. La conferenza ha riunito oltre 150 giovani provenienti da 20 Paesi, offrendo uno spazio di espressione e confronto sul ruolo delle nuove generazioni nel contrasto allo sviluppo dei robot killer (SKR, 2020).

A questa prima iniziativa, negli anni successivi, si sono aggiunte ulteriori esperienze di coinvolgimento giovanile. Ricordiamo, ad esempio, la simulazione del Primo Comitato dell'Assemblea Generale delle Nazioni Unite, interamente dedicata ai *Lethal Autonomous Weapons Systems* (*LAWS*), realizzata nell'ambito del *World Federation of UN Associations International Model United Nations* (*WIMUN*) di New York nel 2024 (WFUNA, 2024). L'iniziativa ha offerto agli studenti partecipanti l'opportunità di confrontarsi con le implicazioni giuridiche, etiche e strategiche legate all'impiego delle armi autonome, promuovendo una riflessione informata e multilaterale sulle relative prospettive normative.

In una prospettiva complementare, orientata alla divulgazione informale e all'incontro con il pubblico generalista, si collocano altre iniziative come la partecipazione alla Campagna di *SKR* alla *Privacy Week* di Milano, il 29 settembre 2022. In tale occasione, sono state evidenziate le criticità etiche, tecniche e giuridiche associate all'automazione letale, con particolare enfasi sull'importanza del mantenimento del controllo umano sulle decisioni di vita o di morte (info.nodes, 2025).

Infine, il 20 settembre 2024, durante il DIG Festival, le associazioni italiane *DIG* - Documentari Inchieste Giornalismi e info.nodes hanno promosso un momento di divulgazione informale attraverso un aperitivo tematico intitolato *Stop Killer Robots. Quando la guerra la combatte l'intelligenza artificiale*. L'iniziativa ha rappresentato un'occasione di incontro e scambio tra cittadini, attivisti e organizzazioni interessate al tema, con l'intento di avvicinare nuovi interlocutori e rendere la riflessione sulle armi autonome più inclusiva e partecipata (DIG, 2024).

6.4. Comunicazione pubblica

Nel contesto della mobilitazione civile, la comunicazione rappresenta uno spazio strategico in cui visione politica e immaginazione collettiva si incontrano. Non si tratta solo di trasmettere messaggi, ma di modellare percezioni, attivare emozioni e generare risonanze culturali. Diverse modalità espressive contribuiscono così a rafforzare l'impatto pubblico del movimento.

6.4.1. Strategie comunicative e narrazioni mediatiche

Accanto all'azione di *advocacy* e alla produzione di conoscenza, la comunicazione pubblica ha rappresentato una leva fondamentale, mantenere alta l'attenzione internazionale e coinvolgere nuovi soggetti nella riflessione critica sulle armi autonome.

Tra le iniziative internazionali, si segnala la newsletter mensile *The Autonomous Weapons Newsletter* del Future of Life Institute, che offre un monitoraggio aggiornato sugli sviluppi politici e tecnologici legati alla IA militare (Future of Life Institute, 2024b). Anche organizzazioni come Human Rights Watch hanno fatto della comunicazione pubblica uno strumento strategico. In Francia, la sezione nazionale ha diffuso una serie di comunicati volti a rafforzare il sostegno all'adozione di un trattato internazionale. Tra i più recenti, si ricordano: nel gennaio 2024 *Robots tueurs: Le vote à l'ONU devrait ouvrir la voie à la négociation d'un traité d'interdiction* (Human Rights Watch France, 2024a), nel maggio *Un large soutien international en faveur d'un traité sur les 'robots tueurs'* (Human Rights Watch France, 2024b), e ad agosto *Robots tueurs: Un nouveau rapport de l'ONU appelle à la signature d'un traité d'ici à 2026* (Human Rights Watch France, 2024c). In aggiunta, il gruppo *Automated Decision Research* della Campagna *SKR* invia regolarmente newsletter sulle novità e gli sviluppi dell'autonomia nei sistemi d'arma e in altri settori correlati (SKR, 2021b).

In Italia, la Rete Italiana Pace e Disarmo ha contribuito a questa narrazione attraverso comunicati stampa, come quello del 28 marzo 2025 intitolato *Stop Killer Robots: preoccupazione per la ripresa degli attacchi a Gaza e per l'uso militare della IA* (Rete Italiana Pace e Disarmo, 2025b), pubblicati sul sito ufficiale dell'organizzazione, che ospita anche una raccolta aggiornata di risorse e documenti sul tema (Rete Italiana Pace e Disarmo, 2025c).

La comunicazione è passata anche attraverso la televisione. A partire da gennaio 2025, il canale TV2000 ha trasmesso il programma *Algoretica – Noi e l'intelligenza artificiale*, condotto da Monica Mondo. Nella puntata dell'8 febbraio, dedicata alle guerre del futuro e alle armi autonome letali, sono intervenuti Paolo Benanti (Pontificia Università Gregoriana), Nunzia Ciardi (Agenzia per la cybersicurezza nazionale - ACN), Giampiero Massolo (Istituto per gli Studi di Politica Internazionale - ISPI) e Fabrizio Battistelli (Archivio Disarmo), contribuendo alla divulgazione presso il pubblico generalista dei temi cari alla Campagna *SKR* (Borghi, 2025). Tra gli altri programmi televisivi che hanno trattato la IA militare, citiamo il *TG3 Mondo* su Rai3 nel 2020, *Checkpoint* su Rainews 24 nel 2021 e *Today* su TV2000 nel 2023 (Rete Italiana Pace e Disarmo, 2025c).

Una parte rilevante della comunicazione è passata anche sui canali digitali e social media. Sia la Rete Italiana Pace e Disarmo sia Archivio Disarmo hanno promosso le attività della Campagna *SKR* attraverso Facebook, Instagram, Twitter/X e i rispettivi siti web. Tali attività hanno incluso notizie, appelli all'azione, materiali di approfondimento e aggiornamenti regolari, contribuendo alla diffusione virale dei contenuti e alla costruzione di una comunità informata e attiva.

6.4.2. Linguaggi artistici, audiovisivi e premi tematici

Oltre all'impegno sul piano accademico, istituzionale e della sensibilizzazione pubblica, tra il 2020 e il 2025 la società civile ha fatto ricorso anche a strumenti comunicativi innovativi e linguaggi artistici, con l'obiettivo di ampliare il coinvolgimento dell'opinione pubblica e di rafforzare la risonanza culturale dell'impegno contro i robot killer. La produzione di contenuti audiovisivi, l'organizzazione di mostre, la promozione di concorsi creativi e l'assegnazione di premi internazionali hanno rappresentato momenti chiave per raccontare, in modo accessibile e impattante, la sfida posta dallo sviluppo della IA militare.

Nel campo delle arti visive, la mostra *Automated by Design*, curata nel 2023 da *Identity 2.0* in collaborazione con la Campagna *SKR*, Soka Gakkai International e Amnesty International, ha proposto un'esplorazione critica del fenomeno della deumanizzazione digitale. L'esposizione, pensata come esperienza sia fisica sia digitale ha analizzato come l'automazione dei processi decisionali, privi di *CUS*, possa compromettere la dignità e i diritti fondamentali delle persone (SKR, 2023). In quest'ottica, la mostra si inserisce nel più ampio impegno della Campagna *SKR* per contrastare l'uso disumanizzante della IA militare.

Nell'ambito cinematografico, il documentario *Immoral Code*, rilasciato il 24 maggio 2022 dalla stessa Campagna, ha interrogato il pubblico sull'accettabilità morale di affidare a una macchina il potere di decidere della vita e della morte (SKR, 2022). In 23 minuti, il film evidenzia le implicazioni etiche, sociali e politiche delle armi autonome, sottolineando l'urgenza di una regolamentazione internazionale. Il documentario è stato anche presentato in Italia durante il DIG Festival di Modena il 25 settembre 2022, in un panel che ha coinvolto esperti e attivisti, tra cui Davide Del Monte (info.nodes), Catherine Connolly (*SKR*), Barbara Gallo (Archivio Disarmo) e Philip Di Salvo (DIG) (info.nodes, 2025).

Altri soggetti attivi sul fronte della regolazione tecnologica hanno contribuito con contenuti audiovisivi. In particolare, nel 2021 il Future of Life Institute ha prodotto una serie di video volti a sensibilizzare il pubblico sulla minaccia crescente della IA militare. Dopo il cortometraggio *Slaughterbots* del 2017 (Future of Life Institute, 2017), nel 2021 l'organizzazione no-profit ha rilasciato *Slaughterbots - If Human: Kill ( )* (sic), una distopia visiva pensata per illustrare i rischi derivanti da una mancata regolamentazione delle armi autonome (Future of Life Institute, 2021). Infine, nel marzo 2025 è stato presentato l'ultimo cortometraggio della serie, dal titolo *Slaughterbots and the Urgent Fight to Stop Them* (Future of Life Institute, 2025).

Alla dimensione creativa si è affiancata quella del riconoscimento pubblico. Il 12 ottobre 2024, Archivio Disarmo ha conferito alla Campagna *SKR* il *Premio internazionale Colombe d'Oro per la Pace* nel corso della cerimonia tenutasi in Campidoglio, a Roma (IRIAD, 2024b). Ha ritirato il premio il professor Peter Asaro vicepresidente dell'*International Committee for Robot Arms Control* (*ICRAC*). In precedenza, la Campagna era stata insignita anche dello *Ypres Peace Prize* nel 2020 (Human Rights Watch, 2020) ed era stata candidata al Premio Nobel per la Pace nel 2021 dal parlamentare norvegese Audun Lysbakken (Aftenposten, 2021), in riconoscimento del suo ruolo centrale nella promozione del disarmo etico e della tutela della dignità umana.

In linea con questa visione culturale e partecipativa, tra il 21 settembre e il primo dicembre 2024 la Campagna *SKR* ha promosso il concorso internazionale *Future 2045*, rivolto a giovani artisti e creativi. L'iniziativa ha invitato i partecipanti a immaginare un futuro possibile in relazione alle scelte globali sull'uso della IA in ambito bellico. Le opere selezionate da una giuria internazionale sono state presentate all'inizio del 2025 in una mostra digitale e durante il meeting globale della Campagna (SKR, 2024c).

6.5. Osservazioni conclusive

La ricostruzione delle principali iniziative promosse nel quinquennio 2020-2025 ha messo in luce il ruolo della società civile nel dibattito sull'impiego degli *AWS*. In un contesto globale in cui l'innovazione tecnologica procede precipitosamente, la voce della società civile rappresenta un elemento di equilibrio fondamentale, capace di orientare le scelte politiche e normative verso soluzioni più etiche e responsabili.

È proprio questo approccio *bottom-up* che si dimostra essenziale per favorire un cambiamento efficace. Non si tratta soltanto di influenzare i processi decisionali dall'alto, ma anche di costruire consapevolezza collettiva "dal basso verso il basso", ovvero generare sensibilizzazione capillare tra i cittadini comuni, nelle scuole, nei quartieri, nelle parrocchie, nei contesti locali in generale. È in questi spazi che può germogliare una cultura di responsabilità critica nei confronti dell'automazione letale.

Al contempo, crescono gradualmente l'attenzione del pubblico e una maggiore diffusione della conoscenza relativa agli *AWS*. Il dibattito non è più confinato agli ambienti accademici o militari: si moltiplicano i momenti di mobilitazione, le iniziative civiche, gli interventi degli esperti nei media. Questo fermento testimonia una sensibilità crescente nei confronti dei rischi etici e sociali che accompagnano la delega di decisioni di vita o di morte alle macchine.

L'ampio ventaglio di iniziative dimostra che la società civile può fermare o riorientare progetti ad alto impatto e rivela anche una tensione sempre più forte verso forme di giustizia tecnologica. In un tempo in cui la distanza tra chi decide e chi subisce le decisioni tende ad ampliarsi, riaffermare la centralità del giudizio umano, della responsabilità e del controllo democratico è una sfida imprescindibile.

Infine, questa dinamica conferma l'importanza della società civile in relazione all'opinione pubblica: è proprio attraverso l'informazione, il confronto interpersonale e la pressione collettiva che diventa possibile incidere concretamente sulle scelte politiche e normative.

Riferimenti bibliografici

Aftenposten. (2021). *Flere fredsprisforslag før fristen gikk ut*. Disponibile a: https://www.aftenposten.no/norge/politikk/i/jBL23A/flere-fredsprisforslag-foer-fristen-gikk-ut.

Amnesty International. (2015). *Autonomous Weapons Systems: Five Key Human Rights Issues for Consideration*. London: Amnesty International Publications, 5-28.

Amnesty International Italia. (2021). Amnesty International e Stop Killer Robots: "*Siamo ancora in tempo per fermare le macchine assassine*". Disponibile a: https://www.amnesty.it/amnesty-international-e-stop-killer-robots-siamo-ancora-in-tempo-per-fermare-i-killer-robots/.

Article 36. (2024a). *Why the new UK government should support a treaty on autonomous weapons.* Disponibile a: https://article36.org/wp-content/uploads/2024/07/New-UK-government-AWS-treaty.pdf.

Article 36. (2024b). *Opportunities after the UNGA Resolution on Autonomous Weapons: Moving Toward a New Treaty.* Disponibile a: https://article36.org/updates/opportunities-after-the-unga-resolution-on-autonomous-weapons-moving-toward-a-new-treaty/.

Bannenberg, J., & Groupe de recherche et d'information sur la paix et la sécurité. – GRIP. (2022). *Encadrement des robots tueurs: vers des négociations en dehors de la*

Convention sur certaines armes classiques ? - Groupe de recherche et d'information sur la paix et la sécurité. Disponibile a: https://www.grip.org/encadrement-des-robots-tueurs-vers-des-negociations-en-dehors-de-la-convention-sur-certaines-armes-classiques/.

Bonanata, A. (2025). Prevost spiega la scelta del nome Leone XIV: "Per affrontare le sfide dell'intelligenza artificiale.". *RaiNews*. Disponibile a: https://www.rainews.it/articoli/2025/05/prevost-spiega-la-scelta-del-nome-leone-xiv-per-affrontare-le-sfide-dellintelligenza-artificiale-157d78d1-8b66-4199-a34d-14f12ed56a49.html.

Borghi, R. (2025). Tv2000 con "Algoretica" e Paolo Benanti esplora intelligenza artificiale e implicazioni etiche. *Primaonline*. Disponibile a: https://www.primaonline.it/2025/01/03/429199/algoretica-con-paolo-benanti-su-tv2000-esplora-intelligenza-artificiale-e-implicazioni-etiche/.

Camera dei Deputati. (2024). *Bollettino delle Giunte e delle Commissioni - martedì 29 ottobre 2024*. Camera.it. Disponibile a: https://documenti.camera.it/leg19/resoconti/commissioni/bollettini/html/2024/10/29/indiceGenerale.htm#.

Campisi, T. (2024). *Il Papa: bandire le armi letali autonome, tutelare la dignità umana nell'era delle macchine*. Vatican News. Disponibile a: https://www.vaticannews.va/it/papa/news/2024-07/papa-hiroshima-ai-etica-pace-religioni-dignita-umana-tecnologia.html.

Carrer, L., Del Monte, D., & Signorelli, A. D. (a cura di). (2023). *Man in the loop: ricerca e sviluppo dei sistemi d'arma autonomi in Italia*. info.nodes. Disponibile a: https://www.infonodes.org/man-in-the-loop.

CEC - Consiglio ecumenico delle Chiese. (2022). *"Killer Robots: A Campaign Guide for Churches" now available in six languages*. World Council of Churches. Disponibile a: https://www.oikoumene.org/news/killer-robots-a-campaign-guide-for-churches-now-available-in-six-languages.

CEC - Consiglio ecumenico delle Chiese. (2024a). *Faith communities take a stand against lethal autonomous weapons*. World Council of Churches. Disponibile a: https://www.oikoumene.org/news/faith-communities-take-a-stand-against-lethal-autonomous-weapons.

CEC - Consiglio ecumenico delle Chiese. (2024b). *WCC Executive Committee statement: Yearning for Just Peace in Europe*. World Council of Churches. Disponibile a: https://www.oikoumene.org/resources/documents/wcc-executive-committee-statement-yearning-for-just-peace-in-europe.

Centro de Educación e Investigación para la Paz. (2020). *Riesgos globales y multilateralismo: el impacto de la COVID-19.* Disponibile a: https://ceipaz.org/anuario/anuario-2020/.

Centro de Educación e Investigación para la Paz. (2024). *Oportunidades de paz y lógicas de guerra (2023-2024).* Disponibile a: https://ceipaz.org/anuario-2023-2024/.

CIRC - Comitato Internazionale della Croce Rossa. (2022). *What you need to know about autonomous weapons*. Disponibile a: https://www.icrc.org/en/document/what-you-need-know-about-autonomous-weapons.

DIG - Documentari Inchieste Giornalismi. (2024). *Aperitivo "Stop Killer Robots. Quando la guerra la combatte l'intelligenza artificiale."* DIG Festival. Disponibile a: https://dig-awards.org/eventi/off-2024-aperitivo-stop-killer-robots/.

Dipartimento di Fisica - Università di Pisa. (2024). *Third seminar on "Nuclear Weapons, Disarmament, and Proliferation" - Dipartimento di Fisica*. Università Di Pisa. Disponibile a: https://www.df.unipi.it/en/third-seminar-on-nuclear-weapons-disarmament-and-proliferation/.

Docherty, B. (2025). A Hazard to Human Rights. *Human Rights Watch*. Disponibile a: https://www.hrw.org/report/2025/04/28/hazard-human-rights/autonomous-weapons-systems-and-digital-decision-making.

England J. B. (2016). *Towards Policy Clarity on Autonomous Weapons Systems*. Disponibile a: https://www.orfonline.org/public/uploads/posts/pdf/20230524134402.pdf.

Future of Life Institute. (2017). *Slaughterbots*. YouTube. Disponibile a: https://www.youtube.com/watch?v=HipTO_7mUOw.

Future of Life Institute. (2021). *Slaughterbots - if human: kill()*. YouTube. Disponibile a: https://www.youtube.com/watch?v=9rDo1QxI260.

Future of Life Institute. (2024a). A Diplomat's Guide to Autonomous Weapons Systems. *Future of Life Institute*. Disponibile a: https://futureoflife.org/document/diplomats-guide-to-autonomous-weapons-systems/.

Future of Life Institute. (2024b). *Newsletters - Future of Life Institute*. Disponibile a: https://futureoflife.org/newsletters/.

Future of Life Institute. (2025). *Slaughterbots and the urgent fight to stop them*. YouTube. Disponibile a: https://www.youtube.com/watch?v=Y-ZdmiXbzsE.

Geneva Academy of International Humanitarian Law and Human Rights. (2024). *Sending up a flare: autonomous weapons systems proliferation risks to human rights and international security*. Disponibile a: https://www.geneva-academy.ch/joomlatools-files/docman files/Sending%20Up%20a%20Flare%20Autonomous%20Weapons%20Systems%20Proliferation%20Risks.pdf.

Geneva Academy of International Humanitarian Law and Human Rights, & Comitato Internazionale della Croce Rossa – CICR. (2024). *Expert consultation report on AI and Related Technologies in Military Decision-Making on the Use of Force in Armed Conflicts. Current developments and potential implications*. Disponibile a: https://www.geneva-academy.ch/joomlatools-files/docman-files/Artificial%20Intelligence%20And%20Related%20Technologies%20In%20Military%20Decision-Making.pdf.

GRIP - Groupe de recherche et d'information sur la paix et la sécurité. (2021). *Webinaire : "Robots tueurs : la Belgique plaidera-t-elle pour une interdiction ?" (Zoom,*

25 juin, 12h-13h). Disponibile a: https://www.grip.org/event/webinaire-robots-tueurs-la-belgique-plaidera-t-elle-pour-une-interdiction-zoom-25-juin-12h-13h/.

GRIP - Groupe de recherche et d'information sur la paix et la sécurité. (2022). *"L'interdiction des robots tueurs : que peut faire la Belgique ?" Webinaire GRIP – 15 mars 2022 (12h30)*. Disponibile a: https://www.grip.org/linterdiction-des-robots-tueurs-que-peut-faire-la-belgique-webinaire-grip-15-mars-2022-12h30/.

Human Rights Watch & International Human Rights Clinic of Harvard Law School. (2012). *Losing Humanity: The Case Against Killer Robots*. Disponibile a: https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots.

Human Rights Watch France. (2024a). *Robots tueurs : Le vote à l'ONU devrait ouvrir la voie à la négociation d'un traité d'interdiction*. Human Rights Watch. Disponibile a: https://www.hrw.org/fr/news/2024/01/03/robots-tueurs-le-vote-lonu-devrait-ouvrir-la-voie-la-negociation-dun-traite.

Human Rights Watch France. (2024b). *Un large soutien international en faveur d'un traité sur les "robots tueurs."* Human Rights Watch. Disponibile a: https://www.hrw.org/fr/news/2024/05/06/un-large-soutien-international-en-faveur-dun-traite-sur-les-robots-tueurs.

Human Rights Watch France. (2024c). *Robots tueurs : Un nouveau rapport de l'ONU appelle à la signature d'un traité d'ici à 2026*. Human Rights Watch. Disponibile a: https://www.hrw.org/fr/news/2024/08/26/robots-tueurs-un-nouveau-rapport-de-lonu-appelle-la-signature-dun-traite-dici-2026.

Human Rights Watch. (2020). *Children Vote to Stop Killer Robots*. Disponibile a: https://www.hrw.org/news/2020/06/09/children-vote-stop-killer-robots.

info.nodes. (2025). *Stop Killer Robots*. Disponibile a: https://www.infonodes.org/stop-killer-robots.

IRIAD - Istituto di Ricerche Internazionali Archivio Disarmo. (2020). LAWS lethal autonomous weapon systems. La questione delle armi letali autonome e le possibili azioni italiane ed europee per un accordo internazionale. Rapporto di ricerca realizzato con il sostegno del Ministero degli Affari Esteri e della Cooperazione Internazionale. *IRIAD Review. Studi sulla pace e sui conflitti*, *07-08*. Disponibile a: https://www.archiviodisarmo.it/view/K0Y2nX8-UWjKHNM9OQ83o96Kv0-oDTrYQW5IYIxM1dE/iriad-review-luglio-agosto.pdf.

IRIAD - Istituto di Ricerche Internazionali Archivio Disarmo. (2023). *Intelligenza artificiale: pace o guerra?*. Disponibile a: https://www.archiviodisarmo.it/intelligenza-artificiale-pace-o-guerra-comunicato-stampa.html.

IRIAD - Istituto di Ricerche Internazionali Archivio Disarmo. (2024a). *"Intelligenza" delle macchine e follia della guerra: le armi letali autonome*. Disponibile a: https://www.archiviodisarmo.it/intelligenza-delle-macchine-e-follia-della-guerra-le-armi-letali-autonome.html.

IRIAD - Istituto di Ricerche Internazionali Archivio Disarmo. (2024b). *Premio Archivio Disarmo - Colombe d'Oro per la Pace 2024 (XL edizione)*. Disponibile a:

https://archiviodisarmo.it/premio-archivio-disarmo-colombe-d-oro-per-la-pace-2024.html.

Ministero federale austriaco per gli Affari europei e internazionali. (2024). *2024 Vienna Conference on Autonomous Weapons Systems*. Bundesministerium Für Europäische Und Internationale Angelegenheiten Der Republik Österreich. Disponibile a: https://www.bmeia.gv.at/en/european-foreign-policy/disarmament/conventional-arms/autonomous-weapons-systems/2024-vienna-conference-on-autonomous-weapons-systems.

9Colonne. (2024). *Quirinale, Mattarella riceve delegazione IRIAD*. Disponibile a: https://www.9colonne.it/485286/quirinale-mattarella-riceve-delegazione-iriad.

Public Citizen. (2024). Deadly and Imminent: The Pentagon's Mad Dash for Silicon Valley's AI Weapons. Disponibile a: https://www.citizen.org/article/deadly-and-imminent-report/.

Rancilio, G. (2025). Sull'intelligenza artificiale lo sguardo di Leone XIV. *Avvenire*. Disponibile a: https://www.avvenire.it/rubriche/pagine/sull-intelligenza-artificiale-lo-sguardo-di-leone-xiv.

Rautenbach, P. (2022). On Integrating Artificial Intelligence with Nuclear Control. *Arms Control Association*. Disponibile a: https://www.armscontrol.org/act/2022-09/features/integrating-artificial-intelligence-nuclear-control.

Reaching Critical Will. (2020). *Feminist perspectives on autonomous weapon systems*. Disponibile a: https://reachingcriticalwill.org/resources/publications-and-research/publications/14975-feminist-perspectives-on-autonomous-weapon-systems.

Rete Italiana Pace e Disarmo. (2024). *Giornate di azione contro le spese militari 2025*. Disponibile a: https://retepacedisarmo.org/2024/il-lavoro-congiunto-della-societa-civile-internazionale-per-il-disarmo-umanitario/.

Rete Italiana Pace e Disarmo. (2025a). *L'ABC della PACE: scienza e tecnologia*. Disponibile a: https://retepacedisarmo.org/evento/labc-della-pace-scienza-e-tecnologia/.

Rete Italiana Pace e Disarmo. (2025b). *Stop Killer Robots*. Disponibile a: https://retepacedisarmo.org/stop-killer-robots/.

Rete Italiana Pace e Disarmo. (2025c). *Stop Killer Robots: preoccupazione per la ripresa degli attacchi a Gaza e per l'uso militare della IA - Stop Killer Robots*. Disponibile a: https://retepacedisarmo.org/stop-killer-robots/2025/03/stop-killer-robots-preoccupazione-per-la-ripresa-degli-attacchi-a-gaza-e-per-luso-militare-della-ia.

RUSI - Royal United Services Institute. (2024). *The Proliferation Risk of Lethal Autonomous Weapons – Paper Launch*. Rusi.org. Disponibile a: https://my.rusi.org/events/the-proliferation-risk-of-lethal-autonomous-weapons-paper-launch.html.

SKR - Campagna Internazionale Stop Killer Robots. (2020). *Keeping #YouthInTheLoop*. Disponibile a: https://www.stopkillerrobots.org/news/global-youth-conference-2020/.

SKR - Campagna Internazionale Stop Killer Robots. (2021a). *Stopping Killer Robots: A Guide for Policy Makers*. Disponibile a:

https://www.stopkillerrobots.org/resource/stopping-killer-robots-a-guide-for-policy-makers/.

SKR - Campagna Internazionale Stop Killer Robots. (2021b). *Research and Monitoring*. Disponibile a: https://www.stopkillerrobots.org/research-and-monitoring/.

SKR - Campagna Internazionale Stop Killer Robots. (2022). *Immoral Code*. Disponibile a: https://www.stopkillerrobots.org/take-action/immoral-code/.

SKR - Campagna Internazionale Stop Killer Robots. (2023). *Automated by Design*. Disponibile a: https://automatedbydesign.stopkillerrobots.org.

SKR - Campagna Internazionale Stop Killer Robots. (2024a). *The story so far*. Disponibile a: https://www.stopkillerrobots.org/the-story-so-far.

SKR - Campagna Internazionale Stop Killer Robots. (2024b). *Parliamentary engagement*. Disponibile a: https://www.stopkillerrobots.org/resource/parliamentary-engagement/.

SKR - Campagna Internazionale Stop Killer Robots. (2024c). *Youth Art Contest – Future 2045*. Disponibile a: https://www.stopkillerrobots.org/take-action/youth-art-contest-future-2045/.

SKR - Campagna Internazionale Stop Killer Robots. (2025a). *Our member organisations*. Disponibile a: https://www.stopkillerrobots.org/a-global-push/member-organisations/.

SKR - Campagna Internazionale Stop Killer Robots. (2025b). *List of Parliamentary Pledge Signatories*. Disponibile a: https://www.stopkillerrobots.org/list-of-parliamentary-pledge-signatories/.

UK SKR - Campagna Internazionale Stop Killer Robots Regno Unito. (2022). *Summary of recent UK parliamentary engagement on autonomous weapons and analysis of the UK government's answers to written questions submitted by parliamentarians.* Disponibile a: https://ukstopkillerrobots.org.uk/wp-content/uploads/2022/12/UK-Stop-Killer-Robots-Parliamentary-Summary-and-Analysis30.pdf.

UNA - UK - United Nations Association. (2022). *UK Stop Killer Robots - 2022 Parliamentary Summary and Analysis*. Disponibile a: https://una.org.uk/news/UK-Stop-Killer-Robots-2022-Parliamentary-Summary-and-Analysis.

Università di Oxford. (2024). *Ethics in AI Lunchtime Research Seminar - AI and the Military: Beyond Autonomous Weapons to Strategic Enterprise Transformation | University of Oxford*. Ox.ac.uk. Disponibile a: https://www.ox.ac.uk/event/ethics-ai-lunchtime-research-seminar-ai-and-military-beyond-autonomous-weapons-strategic.

UNSG - United Nations Secretary-General (2024). *Lethal autonomous weapons systems*: Report of the Secretary-General. Disponibile a: https://digitallibrary.un.org/record/4059475?v=pdf.

Vatican News. (2024). *Testo integrale del discorso di Papa Francesco al G7*. Disponibile a: https://www.vaticannews.va/it/papa/news/2024-06/papa-discorso-integrale-g7-puglia-intelligenza-artificiale.html.

Vienna Center for Disarmament and Non-Proliferation. (2024). *Lethal autonomous weapon systems: Where are we and what's next?* Disponibile a: https://vcdnp.org/laws-where-are-we/.

WASP-HS - Wallenberg AI, Autonomous Systems and Software Program - Humanity and Society. (2024). *AI for Humanity and Society 2024 Workshop 3 – In Defense of Dignity in the Face of the Lethal Use of Artificial Intelligence*. Disponibile a: https://wasp-hs.org/ai-for-humanity-and-society-2024-workshop-3-in-defense-of-dignity-in-the-face-of-the-lethal-use-of-artificial-intelligence/.

WFUNA - World Federation of United Nations Associations. (2024). *WIMUN New York 2024*. Disponibile a: https://wfuna.org/program/wimun-new-york-2024/.

WILPF - Women's International League for Peace and Freedom. (2025). *WeAreWILPF: Autonomous weapons and engaging in WILPF's work on disarmament - WILPF*. Disponibile a: https://www.wilpf.org/calendar/wearewilpf-autonomous-weapons-and-engaging-in-wilpfs-work-on-disarmament/.

Appendice 1 - Organizzazioni aderenti alla Campagna *SKR* per Paese NATO/UE

Paesi	**Organizzazioni membri di *SKR***
Austria	Austria Kampagne zum Verbot von Killer Robotern
	ICBL-CMC Austria
	VICESSE (Vienna Centre For Societal Security)
Belgio	Groupe de recherche et d'information sur la paix et la sécurité (GRIP)
	Pax Christi Vlanderren
	Vredesactie
	Vrouwenraad
Canada	Canadian Unitarians for Social Justice
	Canadian Voice of Women for Peace
	Engineers Without Borders Canada
	Manitobans Against Indiscriminate Weapons
	Mines Action Canada
	Project Ploughshares
	Science for Peace
Danimarca	Esbjerg Peace Movement
Finlandia	Aseistakieltäytyjäliitto (the Union of Conscientious Objectors)
	Naiset Rauhan Puolesta (Women for Peace)
	Rauhanliitto (Peace Union of Finland)
	Suomen Rauhanpuolustajat (Finnish Peace Committee)
	Suomen Sadankomitea ry (Committee of 100)
	Tekniikka elämää palvelemaan (Technology for Life)
Francia	Human Rights Watch France
	Initiatives pour le Désarmement Nucléaire (IDN)
	Observatoire des Armaments
	Sciences Citoyennes
Germania	Centre for Feminist Foreign Policy-Germany
	Deutsche Friedensgesellschaft-Vereinigte KriegsdienstgegnerInnen
	Facing Finance
	Love for Life
	Urgewald
	WILPF-Germany
Irlanda	Afri
	Irish Council for Civil Liberties
	Pax Christi Ireland
Islanda	Icelandic Institute for Intelligent Machines
Italia	info.nodes
	Istituto di Ricerche Internazionali Archivio Disarmo (IRIAD/Archivio Disarmo)
	Rete Italiana Pace e Disarmo - Italian Network for Peace and Disarmament
	Unione degli Scienziati Per Il Disarmo ONLUS (Union of Scientists for Disarmament)
Norvegia	Changemaker
	Norges Fredslag (Norwegian Peace Association)
	Norges Fredsråd (Norwegian Peace Council - umbrella of 19 peace organisations)
	Norwegian People's Aid

Paesi Bassi	PAX
	Stichting Vredesburo Eindhoven
Polonia	The Civil Affairs Institute
Regno Unito	Article 36
	Acronym Institute
	Action on Armed Violence
	Amnesty International UK
	Campaign Against Arms Trade
	Campaign for Nuclear Disarmament
	Centre for Feminist Foreign Policy-UK
	Centre for Peace, Security and Armed Violence Prevention
	Drone Wars UK
	International Observatory of Human Rights (IOHR)
	The Methodist Church in Britain
	Scientists for Global Responsibility
	United Nations Association-UK
	War on Want
Slovenia	Slovenian Artificial Intelligence Society
	Slovenian Association for Cognitive Science
Spagna	Asociación Española de Investigación para la Paz (AIPAZ)
	Centre Delàs d'Estudis per la Pau
	Centro de Educación e Investigación para la Paz (CEIPAZ)
	Euskal Herriko Mugarik Gabeko Ingeniaritza
	Fundación Cultura de Paz
	Fundació per la Pau
	Gernika Gogoratuz (Remembering Gernika)
	Instituto de Derechos Humanos, Democracia Cultura de Paz y No Violencia (DEMOSPAZ)
Stati Uniti d'America	Code Pink
	Justice for All
	Human Rights Watch
	Pax Christi Northern California
	Peace Action New York
	Psychologists for Social Responsibility
	Public Citizen
	Surgeons OverSeas
	Upstate Drone Action Coalition
	Washington Office on Latin America
Svezia	Civil Rights Defenders
	Sparvnastet (Stockholm Hackerspace)
	The Swedish Peace and Arbitration Society (SPAS)
	WILPF-Sweden
	www.robotarforfred.se
Ungheria	Állítsuk meg a Gyilkos Robotokat kampány aktivistái

Elaborazione Archivio Disarmo su SKR, 2025*a*

Cap. 7 – L'opinione pubblica italiana e l'autonomia delle armi

7.1. Premessa

Come mostrano le analisi condotte nei precedenti capitoli, il progresso dell'Intelligenza Artificiale (IA) applicata alla sfera militare ha accelerato lo sviluppo di sistemi capaci di identificare, selezionare e colpire bersagli senza l'intervento umano diretto. Nei precedenti capitoli abbiamo visto come questo scenario ponga una sfida radicale ai principi fondanti del *Diritto Internazionale Umanitario* (*DIU*) – in particolare ai criteri di distinzione, proporzionalità e responsabilità giuridica individuale – e come sollevi quesiti sul piano della dignità umana, toccando il nodo della "deumanizzazione algoritmica", ossia il rischio che decisioni di vita o di morte vengano prese su basi probabilistiche e non deliberative (Sparrow, 2007; Asaro, 2012).

Dal punto di vista della politica internazionale, la corsa alle armi autonome rappresenta inoltre un rischio di destabilizzazione sistemica. In assenza di trattati vincolanti, è prevedibile che la tentazione di acquisire un vantaggio competitivo nell'ambito strategico attraverso l'adozione precoce di IA militare inneschi una "spirale di innovazione bellica" non trasparente e fuori dal controllo democratico. D'altro canto, come evidenzia Scharre (2018), la mancanza di supervisione umana sul campo di battaglia può aprire a scenari di escalation automatizzata dei conflitti. A livello tattico, sul campo di battaglia la decisione di colpire verrebbe presa in millisecondi senza un'adeguata valutazione umana della situazione o senza una valutazione umana *tout court*. A livello strategico l'illusione di far combattere le guerre dai "robot" fra loro, e quindi a "perdite zero" per gli eserciti, indebolirebbe nei decisori la riluttanza a ricorrere alla forza, che i cittadini avversano per il timore delle inevitabili vittime.

7.2. L'opinione pubblica internazionale e gli *AWS* tra professionisti e cittadini

L'introduzione delle armi autonome letali suscita dunque controversie, oltre che giuridiche, anche e forse soprattutto etico-politiche, rendendo urgente un dibattito *competente* (in quanto sostenuto da studi scientifici multidisciplinari) e *partecipato* (in quanto coinvolgente l'opinione pubblica dei differenti Paesi). Sarebbe opportuno che tale dibattito fosse caratterizzato da un flusso di comunicazione duplice: dalla politica, dalla comunità scientifica e dai mass media verso la popolazione ma, di converso, pure da parte di quest'ultima, grazie alla libera espressività dei movimenti, delle associazioni e delle istituzioni della società civile. Ciò senza escludere la consultazione professionale, effettuata periodicamente dagli istituti di sondaggi, impiegando le migliori pratiche di rilevazione demoscopica, sulle opinioni dei cittadini.

A livello internazionale il primo attore a esprimersi è stata la comunità scientifica, la quale ha richiamato l'attenzione sulla criticità dello sviluppo e della diffusione delle armi autonome letali. Già nell'ottobre 2013, 272 esperti di informatica, robotica e Intelligenza Artificiale di 37 Paesi, tra cui figure di spicco come Geoffrey Hinton, Alan Bundy e Lucy

Suchman, hanno firmato una dichiarazione chiedendo il bando delle armi che decidono autonomamente di applicare la forza letale, senza controllo umano (ICRAC, 2013). Questa presa di posizione è stata una delle prime e più autorevoli nel dibattito globale.

Due anni dopo, nel 2015, oltre 2.000 ricercatori in IA e robotica hanno firmato una lettera aperta promossa dal *Future of Life Institute*, chiedendo un divieto preventivo della IA militare (Future of Life Institute, 2015). Nel 2023, con lo *Statement on AI Risk* del Center for AI Safety (2023), più di 600 scienziati e accademici hanno ribadito la necessità di introdurre norme giuridiche internazionali vincolanti per proibire l'autonomia letale nei sistemi d'arma (Center for AI Safety, 2023).

Anche le accademie scientifiche dei Paesi del G7 hanno espresso preoccupazione per l'impiego della IA in ambito militare. Il 25 e 26 marzo 2019, riunite a Parigi, hanno adottato la dichiarazione *Artificial Intelligence and Society*, sottolineando la necessità di un uso responsabile e regolamentato di queste tecnologie. Alla redazione del documento ha partecipato anche la delegazione italiana, guidata da Giorgio Parisi, all'epoca presidente dell'Accademia Nazionale dei Lincei. In assenza di una normativa specifica, ogni sistema d'arma dovrebbe comunque rispettare il *DIU*, garantendo trasparenza, chiarezza nelle responsabilità e un *Controllo Umano Significativo* (G7 Science Academies, 2019; Parisi, 2023).

A livello nazionale, diverse posizioni riflettono l'allarme della comunità scientifica. In Belgio, ad esempio, 117 scienziati hanno firmato nel 2017 una lettera aperta per chiedere il divieto nazionale dei robot killer (RTBF, 2017), contribuendo al raggiungimento di una risoluzione parlamentare del 2018 che vieta tali sistemi (PAX, 2018). Nel 2023, il Paese ha confermato il bando diventando uno dei primi Stati al mondo ad adottare una simile misura (The Brussels Times, 2023).

In Italia, nel 2019 oltre 110 ricercatori hanno aderito a un appello promosso da *USPID*, denunciando i rischi etici, sociali ed economici associati allo sviluppo di armi autonome letali (USPID, 2019). In Svizzera, nel 2021, un gruppo di ricercatori ha inviato una lettera al Governo federale per esprimere forte opposizione all'utilizzo di algoritmi in decisioni di vita o di morte (SonntagsBlick, 2021). Appelli simili erano già giunti nel 2017 anche da Australia e Canada, in vista dei negoziati alle Nazioni Unite (The Guardian, 2017).

Analogamente, in un sondaggio su un campione di 524 ricercatori intervistati nel campo della IA e del *machine learning* una significativa maggioranza si oppone all'uso militare dell'Intelligenza Artificiale, manifestando forti preoccupazioni etiche (Zhang *et al.*, 2021).

Infine, citando tra le fonti di crescente preoccupazione l'integrazione della IA nei sistemi militari e nei processi decisionali, il *Bulletin of the Atomic Scientists* (2025) ha spostato nel gennaio 2025 l'Orologio dell'Apocalisse a soli 89 secondi dalla mezzanotte.

Nel complesso, emerge l'immagine di una comunità scientifica acutamente consapevole delle implicazioni etiche e sociali del proprio lavoro, che rivendica una governance più robusta e trasparente della IA e che sottolinea la necessità di coinvolgere attivamente esperti del settore nella definizione di politiche e regolamentazioni che guidino l'uso etico della IA.

Alle prese di posizione di studiosi e ricercatori si affianca un crescente interesse per il punto di vista dei militari stessi, spesso esclusi dal dibattito pubblico ma direttamente coinvolti nell'uso di tecnologie emergenti. Studi condotti in Stati Uniti (Lushenko, 2023), Australia (Galliott *et al.*, 2021) ed Estonia (Pekarev, 2023) evidenziano come la maggioranza dei militari siano d'accordo nel considerare le macchine basate sulla IA come strumenti di supporto alle decisioni. Pur mostrando apertura verso i sistemi autonomi e un riconoscimento di un inevitabile sviluppo di sistemi militari senza equipaggio, anche i militari ribadiscono la necessità del *Controllo Umano Significativo*, auspicando un futuro del campo di battaglia che rimanga centrato sull'intervento umano e sulle responsabilità etiche e professionali di soldati e ufficiali.

Come è naturale, nei confronti della ricerca e sviluppo in materia di armi autonome alle posizioni critiche e alle preoccupazioni che suscitano le imminenti o già realizzate adozioni di queste tecnologie, si contrappongono le valutazioni positive e le rassicurazioni dei settori impegnati nella ricerca e produzione di beni e servizi basati sulla IA a scopi militari. Ad esempio la Fondazione Leonardo-Civiltà delle Macchine, presieduta dall'ex-presidente della Camera dei deputati Luciano Violante, sottolinea che la competizione per la leadership nella IA militare tra le maggiori potenze strategiche (Stati Uniti, Russia e Cina) non corrisponderebbe alla tradizionale definizione di corsa agli armamenti; ciò in quanto la nuova tecnologia è in una fase iniziale di sviluppo, attualmente "lontana dallo stravolgimento del campo militare" (Fondazione Leonardo Civiltà delle Macchine 2022, p. 61, cit. in Battistelli 2024).

Nella convinzione che sia soprattutto l'opposizione dei cittadini allo sviluppo e all'impiego dei *Sistemi d'Arma Autonomi* (*AWS*) a convincere i governi ad adottare norme in grado di prevenire questa minaccia, nel 2017 la Campagna *Stop Killer Robots* (*SKR*) ha lanciato un'ampia indagine *cross-national* volta a rilevare l'atteggiamento dell'opinione pubblica di 23 Paesi nei confronti della IA militare. Tale rilevazione è stata riproposta nel 2018 e nel 2021 arrivando a coinvolgere 28 Paesi. Dal sondaggio del 2021 emerge come in 27 Paesi su 28 (unica eccezione l'India) i cittadini si dichiarino contrari all'impiego della IA in guerra. Altre indagini sperimentali condotte negli ultimi anni hanno dimostrato come l'opinione pubblica internazionale, nella maggior parte dei casi, mantenga una posizione fortemente critica nei confronti delle applicazioni belliche della IA.

La percezione dell'uso militare di questa tecnologia varia in base al contesto politico e alla narrazione dominante nei singoli Paesi. In Cina, ad esempio, prevale un atteggiamento di "tecno-ottimismo", favorito da un maggiore allineamento con le posizioni governative e da una limitata possibilità di dissenso (Rosendorf *et al*., 2024; Qiao-Franco & Bode, 2023). In altri contesti, come gli Stati Uniti, la percezione della minaccia esterna influisce sul sostegno alla IA nel militare: studi su un campione rappresentativo di 2.500 cittadini mostrano che la consapevolezza dello sviluppo di queste tecnologie da parte di altri Paesi può accrescere l'approvazione del loro impiego (Di Giuseppe *et al.*, 2025). Tuttavia, ciò non esclude una disponibilità alla cooperazione

internazionale, suggerendo che il timore di rimanere indietro convive con il desiderio di regole condivise.

Pur a fronte di differenze tra Paesi, si riscontrano convergenze trasversali: l'opinione pubblica tende a posizioni più critiche rispetto ai rispettivi governi, anche in Paesi promotori della deterrenza basata sulla IA come Stati Uniti e Israele. Secondo la rilevazione Ipsos (2021), la crescente opposizione all'impiego della IA in guerra si concentra su tre fattori principali: l'elevata probabilità di errore algoritmico, l'assenza di meccanismi chiari di attribuzione della responsabilità e le implicazioni per la dignità umana. Esperimenti condotti in Stati Uniti, Germania, Brasile e Cina confermano che il consenso verso i sistemi autonomi cresce solo quando essi sono percepiti come più affidabili degli operatori umani. Al contrario, la percezione che possano agire in modo arbitrario alimenta un'opposizione trasversale. A ciò si aggiunge una diffusa preoccupazione per la difficoltà di imputare responsabilità in caso di incidenti letali e per le conseguenze etico-giuridiche derivanti dal trasferimento alle macchine del potere di decidere sulla vita e sulla morte (Rosendorf *et al.*, 2022; 2024).

Tali tendenze critiche non si limitano ai contesti extraeuropei, ma trovano riscontro anche in Italia. Venendo all'opinione pubblica italiana, la sua opposizione all'impiego delle armi autonome sul campo di battaglia (58% di contrari secondo la rilevazione Ipsos 2021) era emersa ancor più nettamente in un sondaggio realizzato due anni prima dall'Istituto di Ricerche Internazionali Archivio Disarmo (IRIAD, 2019), in cui il 69% dei rispondenti si dichiarava contrario. Questi dati confermano come, anche e soprattutto nel contesto italiano, permanga una diffusa preoccupazione verso l'autonomia letale, in linea con le principali criticità già emerse a livello internazionale.

7.3. Il sondaggio di opinione di Archivio Disarmo *Difebarometro* n. 12, 2025

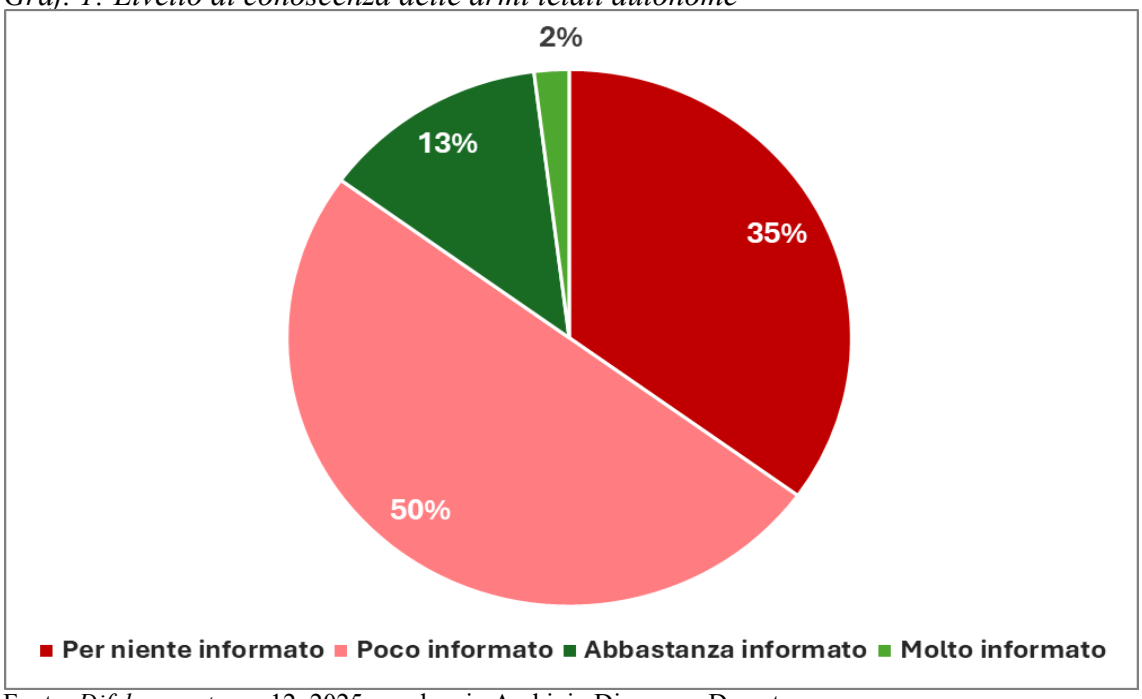
A distanza di alcuni anni dalla nostra prima rivelazione sulle armi autonome, in una fase storica contraddistinta da una svolta critica nell'evoluzione dei conflitti armati e dall'irruzione dell'automazione nei teatri di guerra, nelle pagine che seguono proponiamo l'analisi del sondaggio che Archivio Disarmo ha appositamente effettuato per il presente Rapporto a cavallo tra gennaio e febbraio 2025[1].

[1] Questionario a risposta precodificata, Archivio Disarmo, Difebarometro n. 12, gennaio-febbraio 2025. Rilevazione effettuata da Demetra s.r.l. con metodo CAWI - Computer Assisted Web Interviewing su un campione di 800 rispondenti di età superiore ai 18 anni, rappresentativo della popolazione italiana per genere, età e residenza geografica. Il sondaggio parte della ricerca *Lo stato dell'Intelligenza Artificiale in ambito militare e le prospettive di regolazione a livello nazionale, europeo e internazionale*, commissionato dal Ministero degli Affari Esteri e della Cooperazione Internazionale. I risultati e la loro interpretazione sono responsabilità unica dell'Istituto di Ricerche Internazionali Archivio Disarmo e non rappresentano necessariamente le posizioni del Ministero degli Affari Esteri e della Cooperazione Internazionale.

7.3.1. Un'avversione istintiva? Livello di conoscenza e differenze tra caratteristiche demografiche

Il primo dato che emerge evidenzia una carenza informativa nella popolazione italiana. Dal campo di battaglia ucraino giungono quotidianamente immagini di armi semi-autonome (droni), così come è noto che nel conflitto di Gaza il Governo israeliano impiega sistemi di IA per l'individuazione di soggetti ritenuti membri di Hamas o altre milizie palestinesi; tuttavia, la percentuale di chi dichiara di essere informato – molto o abbastanza – sulle armi letali autonome non supera il 15%, a fronte dell'85% di chi dichiara di essere poco o per niente informato (v. Graf.1).

Graf. 1: Livello di conoscenza delle armi letali autonome



Fonte: *Difebarometro* n. 12, 2025, sondaggio Archivio Disarmo - Demetra

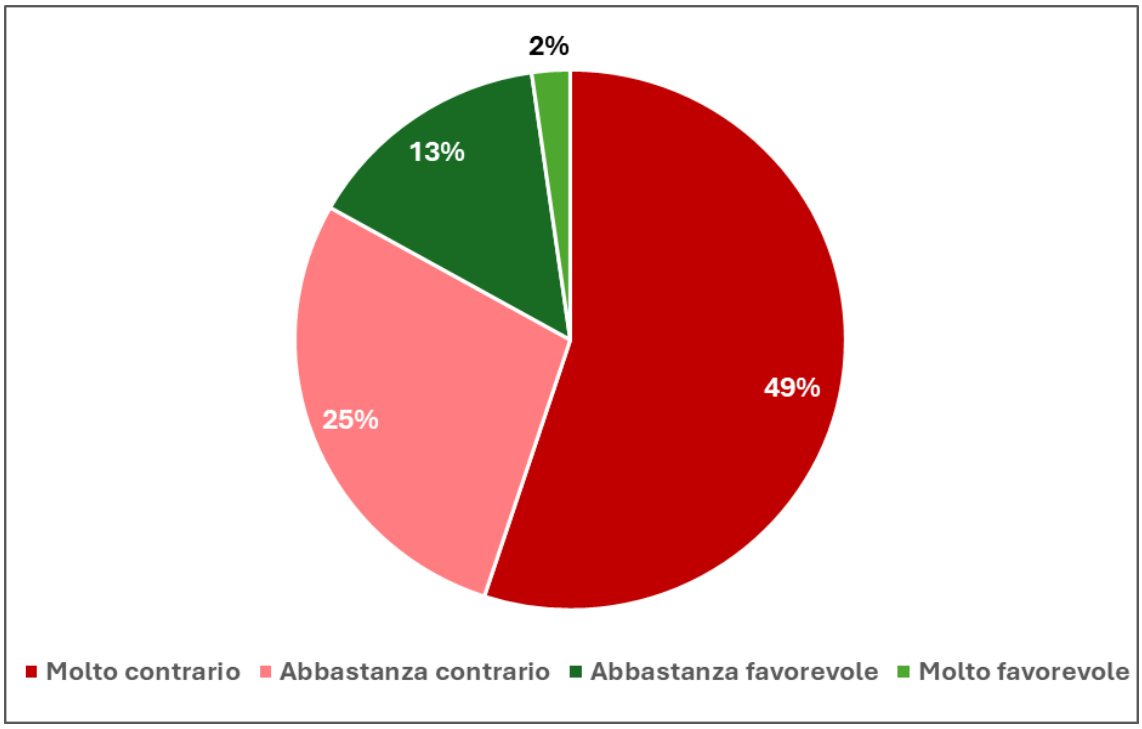
Oltre a rappresentare una conferma di precedenti sondaggi, nostri e di altri, che abbiamo analizzato altrove (Farruggia, 2023), tali risultati non stupiscono se si considera la percezione degli italiani circa le applicazioni della IA in generale, diffusi dall'Osservatorio Scienza Tecnologia e Società - Observa negli ultimi due anni. Infatti, Observa ha rilevato il basso livello di informazione della popolazione italiana rispetto a una tecnologia di cui pure vanno rapidamente diffondendosi le applicazioni, con un 63% dei rispondenti che dichiara di essere poco o per niente informato sulla IA (Bucchi *et al.*, 2024). Nello stesso tempo non si può non sottolineare il differenziale (di ben 22 punti inferiore alla media della conoscenza in generale) che caratterizza la conoscenza della IA in un ambito – quello strategico e militare – dove ancora di più che nel civile gli approfondimenti mediatici sono particolarmente scarsi (v. Graf. 2).

Pur dichiarando un limitato grado di informazione sul tema, ancora una volta l'opinione pubblica italiana si mostra nettamente ostile all'impiego delle armi letali

autonome. Infatti il 74% dei rispondenti si dichiara contrario (nel 49% dei casi *molto contrario*) a tali sistemi d'arma.

Piuttosto che stigmatizzare una simile risposta come meramente istintiva, a questo punto è opportuno esaminare da vicino da chi proviene l'ostilità e sulla base di quali motivazioni.

Graf. 2: Posizione nei confronti delle armi letali autonome



Fonte: *Difebarometro* n. 12, 2025, sondaggio Archivio Disarmo – Demetra

Il primo elemento interessante riguarda la correlazione tra le opinioni espresse dagli intervistati e alcune loro caratteristiche strutturali. Iniziando dal sesso degli intervistati, la contrarietà all'utilizzo delle armi autonome si attesta su livelli molto elevati sia tra gli uomini sia tra le donne, al punto da colmare quel divario di genere (*gender gap*) che solitamente vede il genere femminile meno favorevole all'uso della forza. Si dichiarano infatti molto o abbastanza contrari a tali sistemi il 73% degli uomini e il 74% delle donne, peraltro una differenza talmente ridotta da non risultare statisticamente significativa ($p = 0{,}084$). Più significativo è il fattore età ($p < 0{,}001$): la contrarietà cresce al crescere dell'età anagrafica. Tra i più giovani la contrarietà oscilla tra il 67% dei giovanissimi (18-24) e l'84% degli over 65. Il "pacifismo" dei più anziani è un fenomeno non soltanto italiano bensì si conferma come parte di un orientamento di valore particolarmente diffuso nella coorte demografica di nati nel dopoguerra. Come più volte segnalato in precedenti studi (Caren *et al*., 2011), i *baby boomers* continuano a distinguersi per la marcata diffidenza verso le armi, coerente con l'esperienza della loro generazione, influenzate dalla stagione delle contestazioni politiche e sociali e delle manifestazioni giovanili contro la guerra.

7.3.2. Opinioni "riflessive" e atteggiamenti "affettivi"

Abbiamo inoltre chiesto ai rispondenti di esprimersi relativamente a due quesiti che investono rispettivamente la sfera degli atteggiamenti e quella delle opinioni[2]. Il primo quesito proponeva uno scenario ipotetico in cui l'individuo fosse vittima di un reato commesso da un malintenzionato, e chiedeva di indicare quale comportamento si ritenesse più opportuno: rivolgersi alla polizia, demandando la questione all'autorità pubblica, oppure agire in prima persona per difendersi, anche facendo ricorso a un'arma. Il secondo quesito, invece, chiedeva un'opinione di principio circa l'opportunità di introdurre nel proprio ordinamento la pena di morte per i reati più gravi, come omicidi plurimi o stragi.

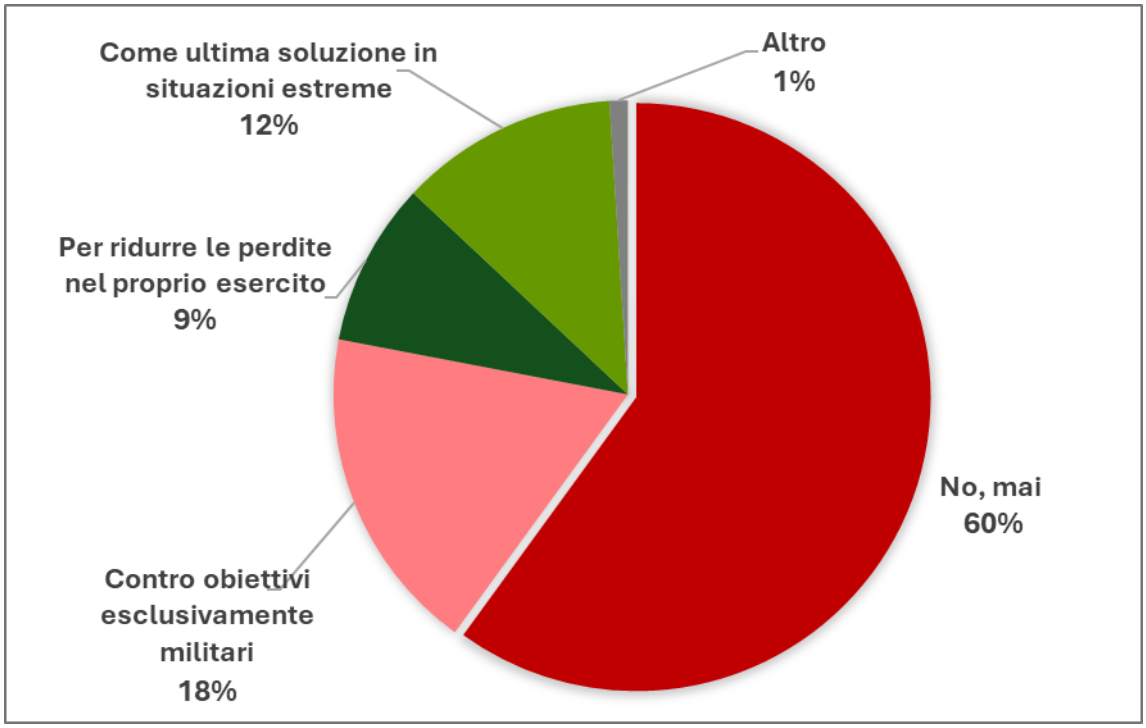
L'analisi incrociata delle risposte a questi due *item* ha consentito di individuare due gruppi idealtipici di rispondenti. Da un lato, coloro che si dichiarano favorevoli sia all'autodifesa armata che alla pena capitale sono stati classificati come "reattivi", espressione di un orientamento punitivo severo e personalizzato. Dall'altro lato, coloro che optano per la tutela istituzionale e che si oppongono alla pena di morte sono stati definiti "adattivi", portatori di una visione più graduata della sanzione, fondata sul principio di legalità e sulla delega allo Stato del monopolio della forza.

I reattivi mostrano una maggiore propensione a favore dell'impiego di armi autonome rispetto ai non reattivi (23% contro 15%). La correlazione tra la reattività individuale e il grado di favore nei confronti di tali tecnologie belliche risulta statisticamente significativa ($p = 0{,}021$), come evidenziato dal test del chi quadrato. Questo risultato suggerisce che un atteggiamento reattivo – inteso come maggiore apertura all'uso della forza in contesti conflittuali – si associa a una più elevata accettazione dell'utilizzo degli *AWS*, delineando un legame rilevante tra predisposizioni psicologiche individuali e giudizi normativi sull'innovazione tecnologica in ambito militare.

È interessante notare che la netta avversione nei confronti dell'impiego di armi letali autonome in guerra emerge con chiarezza anche quando vengono presentate condizioni che potrebbero, teoricamente, giustificarne l'uso: contro obiettivi esclusivamente militari (18%), come *extrema ratio* in situazioni critiche (12%) o al fine di ridurre le perdite tra i propri soldati (9%). Nonostante ciò, la maggioranza assoluta degli intervistati (60%) continua a ritenere che tale impiego non sia mai legittimo.

[2] Nella posizione personale che può essere espressa dai soggetti intervistati in merito a una determinata questione, è opportuno distinguere tra due varianti dell'espressione stessa: quella più immediata e "affettiva" (l'*atteggiamento*) e quella più mediata e "razionale" (l'*opinione*) (Battistelli & Farruggia, 2025).

Graf. 3: Esistono circostanze nelle quali è giustificato l'utilizzo delle armi autonome?



Fonte: *Difebarometro* n. 12, 2025, sondaggio Archivio Disarmo – Demetra

In questo caso si riaffaccia la maggior riluttanza delle donne verso l'utilizzo della forza, occultata nella precedente domanda dalla valanga di no alle armi autonome in quanto tali. Tra chi risponde che le armi autonome non debbano essere utilizzate mai, la maggioranza (56%) è di sesso femminile. Risulta invece confermata la tendenza ad una più netta contrarietà al loro impiego con il crescere dell'età: tra chi risponde che gli *AWS* non dovrebbero essere *mai* impiegati abbiamo ai due estremi i giovanissimi (18-24) che sono solo il 4% dei contrari, mentre gli anziani (over 65) ne rappresentano il 22%.

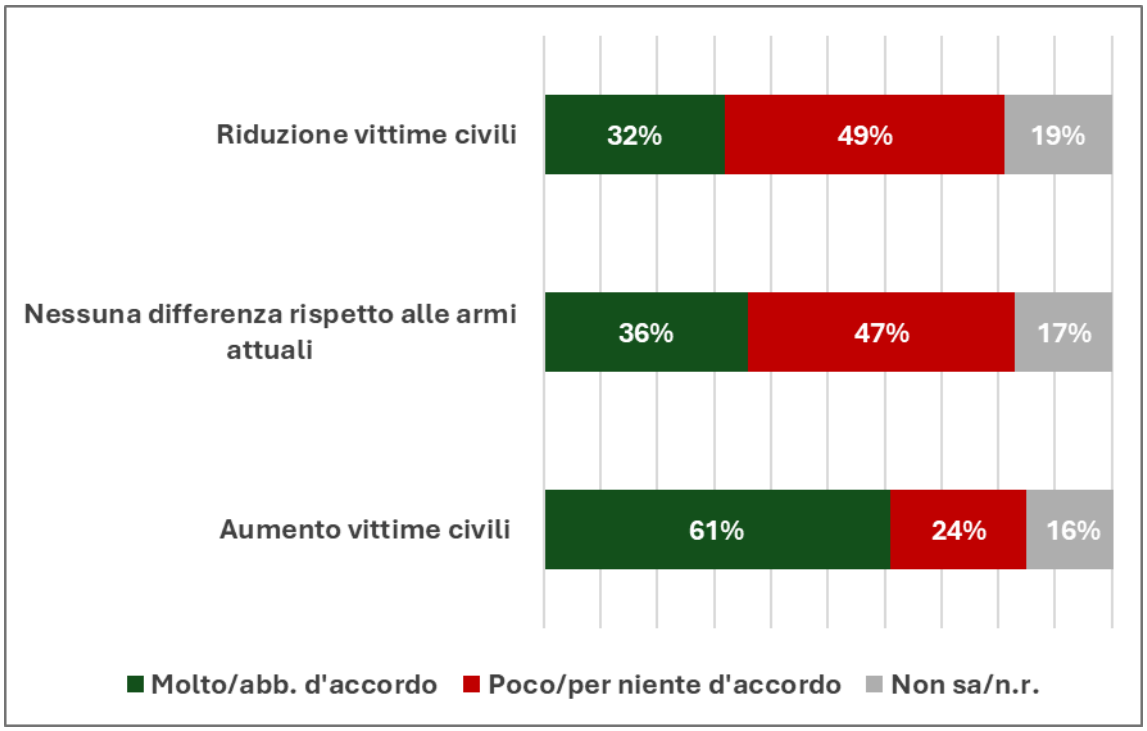
7.3.3. I rischi delle armi autonome

Volendo comprendere le motivazioni che portano a questa netta contrarietà della popolazione italiana nei confronti degli *AWS*, abbiamo dapprima indagato la percezione dei possibili effetti del loro impiego su una delle conseguenze dei conflitti maggiormente avversata dall'opinione pubblica: le vittime civili. Più specificamente abbiamo chiesto agli intervistati se, a loro avviso, l'introduzione di questa nuova tecnologia porterà o meno una riduzione degli "effetti collaterali" delle guerre.

Sul possibile aumento delle perdite non intenzionali causate della IA gli esperti internazionali hanno opinioni divergenti: da un lato vi sono studiosi che sottolineano la maggiore precisione chirurgica di un'arma autonoma (Arkin, 2009), dall'altra c'è chi sostiene che tali livelli di precisione sono ancora lontani dall'essere raggiunti e che, ad ogni modo, la maggiore facilità con cui i governi saranno portati a fare la guerra – tentati dalla riduzione delle perdite tra le proprie truppe – a lungo termine favorirà un più

frequente ricorso ad esse e il conseguente aumento delle vittime tra i civili (Altman, 2013). A tal riguardo, solo il 32% dei rispondenti si dichiara molto o abbastanza d'accordo con l'ipotesi di una possibile riduzione di vittime civili favorita dall'impiego degli *AWS*; di contro al 61% degli intervistati che si dichiara molto o abbastanza d'accordo con l'affermazione che l'uso di armi autonome comporterà un aumento delle vittime civili. Rispetto all'affermazione che tale evoluzione tecnologica non avrebbe alcun impatto – né positivo né negativo – sul numero di vittime civili, si dichiara d'accordo il 36% dei rispondenti (v. Graf. 4).

Graf.4: Possibili effetti dell'impiego delle armi letali autonome (relativamente al numero delle vittime civili)



Fonte: *Difebarometro* n. 12, 2025, sondaggio Archivio Disarmo - Demetra

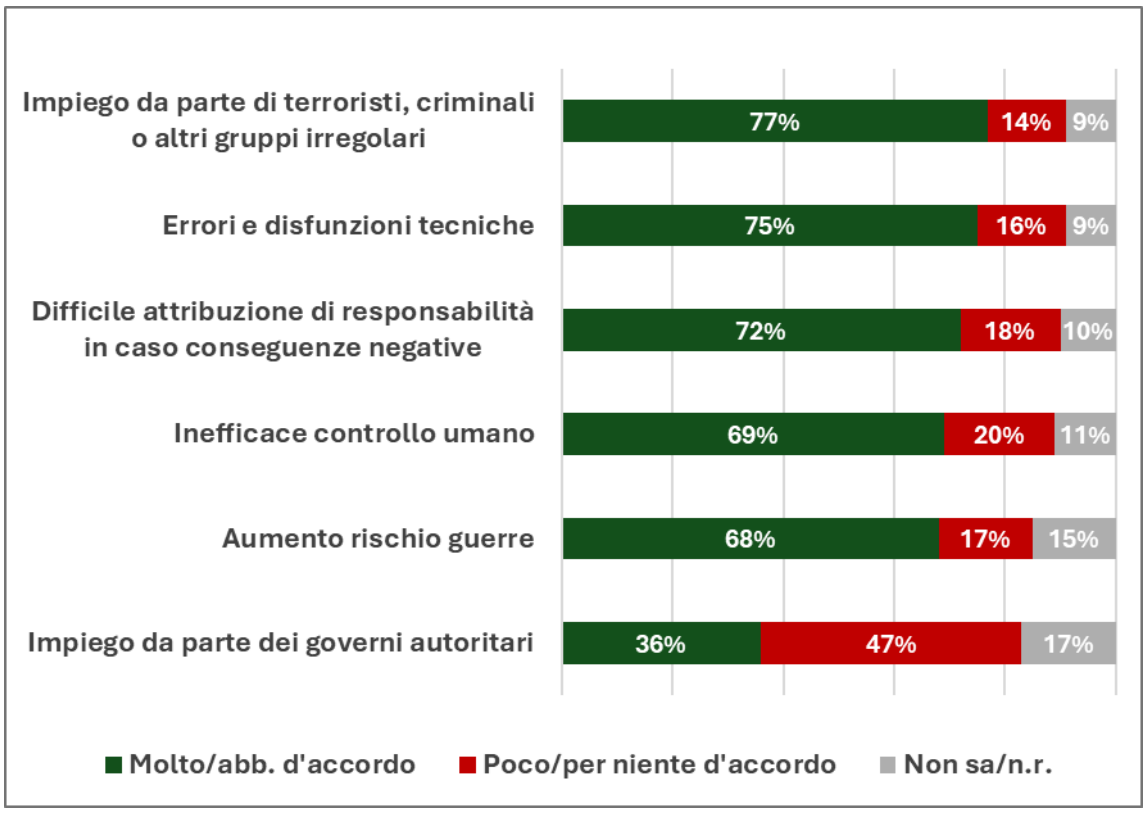
L'opinione pubblica italiana tende dunque a diffidare delle armi autonome in termini di impatto umanitario. L'ipotesi di una riduzione delle vittime convince 1/3 scarso del campione, mentre prevale la convinzione che tali sistemi aumenteranno il rischio per i civili, riflettendo timori etici e pratici circa il venire meno del controllo umano diretto.

Oltre al possibile incremento delle vittime civili, abbiamo posto all'attenzione dei rispondenti gli ulteriori rischi che potrebbero scaturire dall'utilizzo delle armi autonome. Tra gli scenari di rischio prospettati, quello che viene considerato con maggiore preoccupazione dall'opinione pubblica riguarda la possibilità che le armi autonome letali possano essere utilizzate da attori non statali, come gruppi terroristici, organizzazioni criminali o formazioni paramilitari irregolari. Ben il 77% dei rispondenti giudica tale evenienza "molto" o "abbastanza" probabile, rendendola la minaccia percepita più concreta e immediata. Subito dopo, con una quota simile (75%), si colloca il timore di

malfunzionamenti tecnici o errori sistemici, che potrebbero compromettere gravemente l'affidabilità e la sicurezza di questi sistemi. A seguire, un livello elevato di preoccupazione è anche suscitato dall'ardua (o impossibile) attribuzione di responsabilità in caso di incidenti o conseguenze fatali (72%), una problematica giuridica e morale che, come è stato approfondito in precedenza (v. cap. 5), rappresenta uno dei nodi centrali nel dibattito internazionale.

Altri rischi percepiti come rilevanti includono il deficit di controllo umano (69%) e l'eventualità che la disponibilità degli *AWS* possa essere impiegata illecitamente da terroristi o criminali (77%); oppure incorrere in errori o disfunzioni (75%); o rendere impossibile l'attribuzione delle responsabilità (72%); e, infine, con una soglia dell'uso della forza più bassa facilitando le guerre (68%). Relativamente meno incombente lo scenario di un loro impiego da parte di governi autoritari, giudicata "poco" o "per niente" probabile dal 47%. Quest'ultimo dato può essere letto come una sottovalutazione del rischio sistemico insito in contesti politico-istituzionali scarsamente trasparenti, oppure come il riflesso di un *frame* geopolitico attualmente centrato sulle minacce asimmetriche più che su quelle statali.

Graf. 5: Possibili effetti dell'impiego delle armi letali autonome



Fonte: *Difebarometro* n. 12, 2025, sondaggio Archivio Disarmo – Demetra

7.3.4. Responsabilità e controllo da parte umana

In generale, i risultati rivelano un dato interessante: a fronte di una competenza tecnica, ammessa come scarsa, sulla IA militare, la maggioranza dei rispondenti manifesta una percezione dei rischi sostanzialmente in linea con le preoccupazioni espresse da numerosi esperti e scienziati a livello internazionale (Altmann, 2013; Altmann & Sauer, 2017). Questo sembra indicare un'intuizione etico-politica diffusa, che percepisce la natura trasformativa e potenzialmente destabilizzante della IA applicata alla guerra.

Particolare attenzione, in quanto coinvolge direttamente la cruciale discussione sulla legalità delle armi autonome, è il tema della responsabilità nell'uso delle medesime. Come si è visto, quasi 3/4 degli intervistati temono che l'introduzione di sistemi in grado di operare in autonomia possa minare la capacità di attribuire con chiarezza la responsabilità per eventuali violazioni del *DIU*, soprattutto in caso di vittime civili o di condotte considerate illegali. La larga maggioranza dei rispondenti esprime una preoccupazione marcata circa l'impatto che l'impiego di questi sistemi potrebbe avere sul principio di responsabilità individuale, considerato uno dei pilastri del *DIU* (Tamburrini, 2020). A questo proposito, è stato chiesto agli intervistati di indicare chi, a loro avviso, dovrebbe essere ritenuto responsabile nel caso in cui un'arma autonoma provocasse un'uccisione ingiustificata.

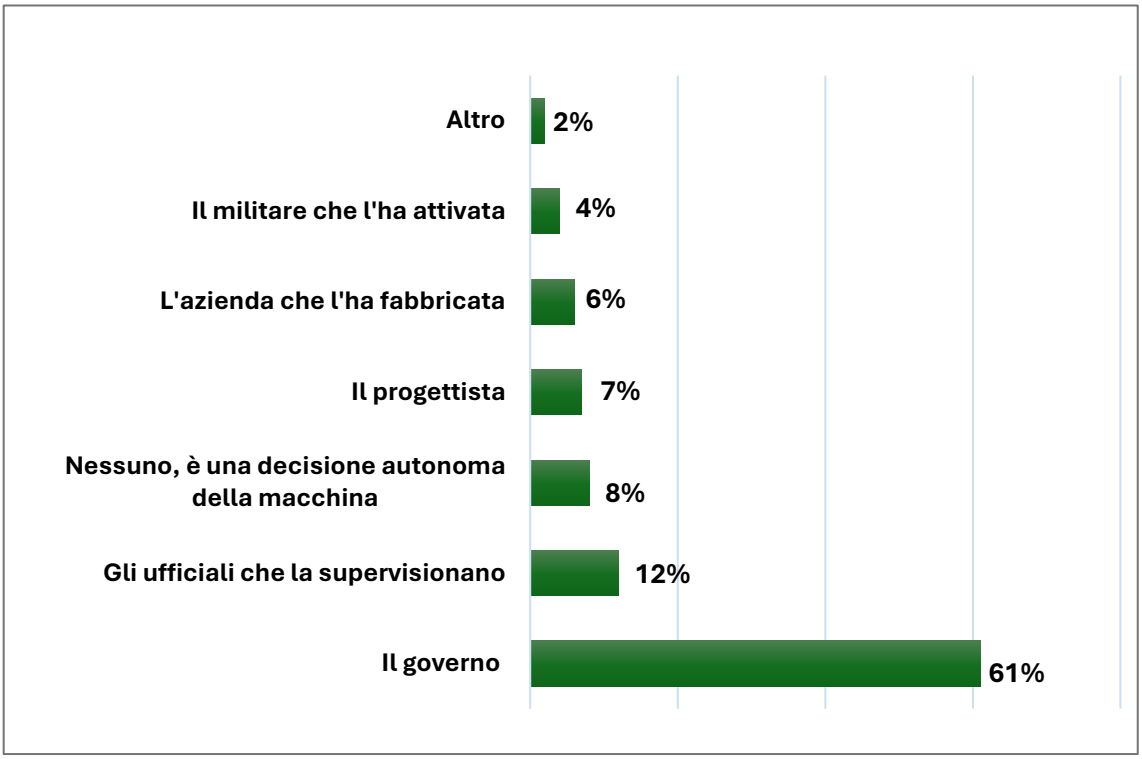
I risultati mostrano una tendenza interpretativa netta: il 61% del campione attribuisce la responsabilità al governo dello Stato che ha impiegato il sistema d'arma, mentre il restante 39% si distribuisce fra altre opzioni: il comandante o ufficiale che supervisiona il sistema (12%), la macchina stessa (8%), il progettista (7%), l'azienda produttrice (6%) e, infine, il militare che ha attivato il dispositivo (4%); altro (2%). Questo orientamento generale suggerisce che l'opinione pubblica tende a privilegiare un approccio politico-istituzionale alla responsabilità, piuttosto che uno strettamente individuale o tanto meno tecnico (v. Graf.6).

È da osservare che questa percezione si pone in controtendenza rispetto all'orientamento giuridico propugnato dal *DIU*, secondo cui – nelle guerre combattute con armamenti convenzionali – la responsabilità non ricade direttamente sugli Stati, bensì sugli ufficiali comandanti, a condizione che essi siano consapevoli delle azioni compiute dai soldati sotto il loro comando (CICR, 2024). L'idea che una macchina possa sostituirsi alla deliberazione umana nel processo di attacco solleva dunque interrogativi cruciali sulla capacità stessa di mantenere in vita il modello attuale di imputazione della responsabilità, fondato sulla coscienza morale e sulla deliberazione individuale.

La consapevolezza diffusa rispetto ai rischi giuridici, morali e strategici legati all'impiego degli *AWS* ha infatti alimentato negli ultimi anni un acceso dibattito sulla necessità di regolamentare queste tecnologie prima della loro piena diffusione sul campo di battaglia. Tuttavia, già nelle fasi iniziali del confronto internazionale è emersa una difficoltà di fondo: l'assenza di una definizione condivisa degli *AWS*, in riferimento ai quali Taddeo e Blanchard (2022) hanno censito ben 14 definizioni differenti. La difficoltà di una definizione univoca rappresenta ancora oggi uno degli ostacoli principali

all'elaborazione di un quadro regolatorio efficace. Da qui l'incertezza su quale tipo di controllo umano debba essere garantito per rispettare i principi del *DIU* (Tamburrini, 2023; v. cap. 1).

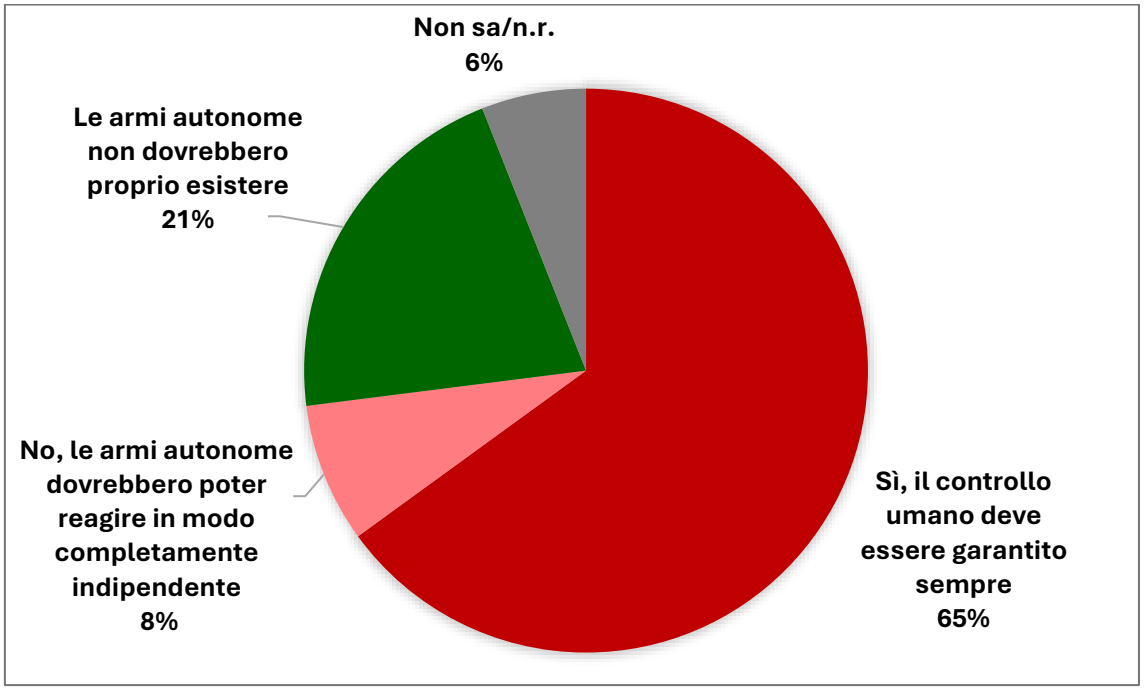
Graf. 6: Soggetto responsabile in caso di uccisioni ingiustificate



Fonte: *Difebarometro* n. 12, 2025, sondaggio Archivio Disarmo – Demetra

L'ultima parte dell'indagine ha affrontato direttamente questo nodo, chiedendo ai rispondenti di esprimersi sul grado di autonomia operativa che tali sistemi d'arma dovrebbero possedere. Coerentemente con gli orientamenti già espressi nei quesiti precedenti, la maggioranza assoluta (65%) ritiene che il controllo umano debba essere sempre garantito, affermando implicitamente il principio di *Controllo Umano Significativo* come criterio minimo di legittimità. Solo una ridotta minoranza (8%) sostiene che le armi autonome dovrebbero poter agire in completa indipendenza, mentre sul fronte opposto una minoranza più corposa (21%) si dichiara contraria alla loro esistenza in qualunque forma, indipendentemente dal grado di autonomia tecnica (v. Graf. 7). Questo dato rafforza l'idea che, al di là della dimensione funzionale, l'applicazione dell'autonomia delle armi continua a essere percepita come una questione eminentemente etico-politica, che chiama in causa i fondamenti stessi del diritto umanitario e della convivenza internazionale.

Graf.7: Ritiene necessario mantenere un controllo umano sulle decisioni di attacco delle armi letali autonome?



Fonte: *Difebarometro* n. 12, 2025, sondaggio Archivio Disarmo - Demetra

7.3.5. Dal controllo tattico all'Arms Control politico e diplomatico

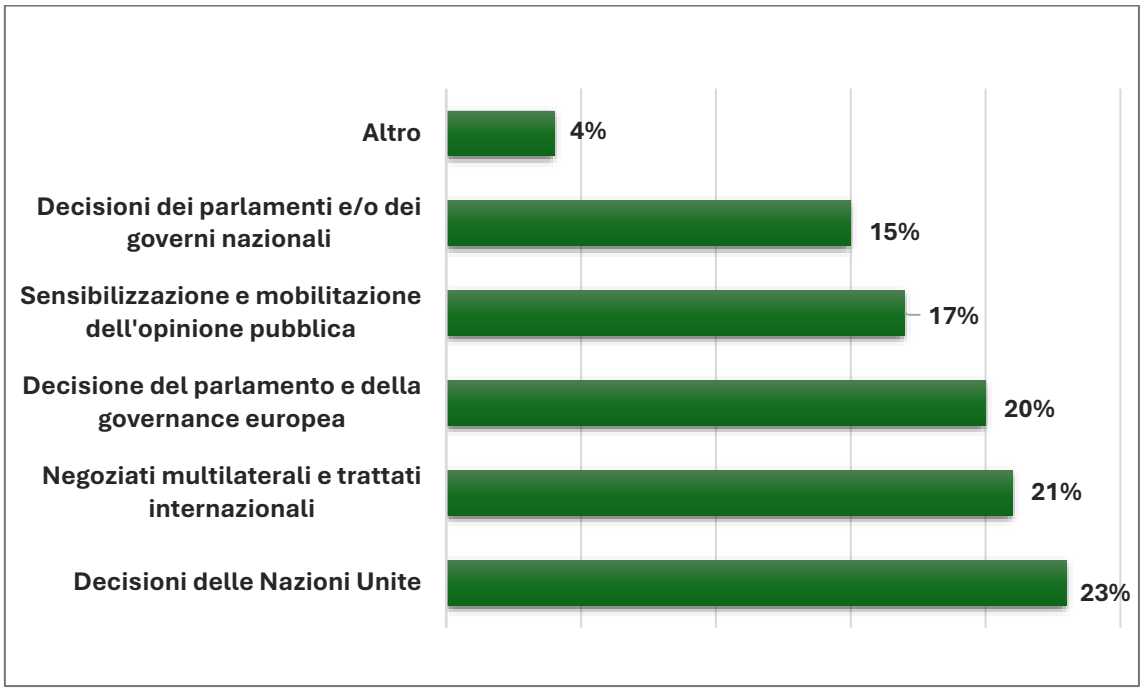
La necessità di garantire un *Controllo Umano Significativo* sull'impiego delle armi autonome sul campo di battaglia emerge dunque come un orientamento condiviso da una maggioranza di 2/3 dell'opinione pubblica italiana. La soluzione appare in tutta la sua complessità quando si passa alla definizione di un percorso istituzionale a livello strategico, volendo affrontare la questione in termini di controllo degli armamenti e prevenzione della loro proliferazione.

Alla domanda su quale dovrebbe essere il canale più adeguato per regolamentare la ricerca, lo sviluppo e l'utilizzo degli *AWS*, le risposte del campione risultano piuttosto frammentate, riflettendo probabilmente la complessità del tema su cui pronunciarsi, oltre a prendere atto della molteplicità degli attori coinvolti a livello nazionale e internazionale.

Nello specifico, solo il 15% degli intervistati ripone fiducia nei parlamenti o nei governi nazionali, a conferma di una diffusa percezione di inadeguatezza dei singoli Stati nel fronteggiare sfide tecnologiche e strategiche di scala globale. Una quota leggermente superiore (17%) attribuisce invece un ruolo centrale alla mobilitazione dell'opinione pubblica, riconoscendo nella pressione dal basso una possibile leva di cambiamento politico e normativo. Il 20% auspica un intervento più strutturato da parte delle istituzioni europee, segno di una fiducia non trascurabile nel potenziale regolativo della governance UE, mentre il 21% individua nei negoziati multilaterali e nei trattati internazionali la via maestra per una normazione efficace. Infine, la quota relativamente più numerosa (23%) si affida alle Nazioni Unite, ribadendo una fiducia nel modello multilaterale tutt'altro che

scontata in un clima ad esso avverso e in una crisi globale delle relazioni internazionali di grandi proporzioni come l'attuale (v. Graf. 8).

Graf.8: Quale sarebbe la misura più efficace per regolamentare la ricerca e l'impiego delle armi autonome?



Fonte: *Difebarometro* n. 12, 2025, sondaggio Archivio Disarmo - Demetra

Questo quadro eterogeneo, pur segnalando l'assenza di un'unica istituzione riconosciuta come legittima ed efficace istanza deliberativa della comunità internazionale, porta comunque alla luce un dato rilevante e non scontato: la somma delle risposte che fanno riferimento a soluzioni multilaterali e sovranazionali – tra Nazioni Unite, Unione Europea e trattati internazionali – raggiunge complessivamente il 64%. Ciò dimostra che, malgrado il crescente scetticismo verso le architetture globali di governance, manifestato da alcuni settori della politica nazionale ed europea, un'ampia maggioranza dell'opinione pubblica italiana continua ad attribuire valore all'approccio del multilateralismo, ritenendolo ancora oggi uno strumento imprescindibile per la gestione condivisa delle sfide globali.

7.4. Osservazioni conclusive

I risultati emersi dal sondaggio effettuato da Archivio Disarmo nel gennaio/febbraio 2025 rilevano con chiarezza la posizione dell'opinione pubblica italiana nei confronti delle armi letali autonome: un atteggiamento di avversione e una domanda di precauzione, basata su convinzioni di natura sia etica sia politica. Nonostante il basso livello di informazione dichiarato dai rispondenti, il quadro di giudizio che emerge è sorprendentemente consapevole nel merito dei problemi, in linea con le principali

preoccupazioni espresse dalla comunità scientifica, dalle organizzazioni internazionali e dalle associazioni della società civile.

In primo luogo si delinea una forte rivendicazione del principio di responsabilità, riservato agli umani, peraltro uno dei pilastri irrinunciabili del *DIU*. Il timore che l'autonomia decisionale delle macchine possa rendere difficile – se non impossibile – l'attribuzione di colpa in caso di violazioni gravi del diritto di guerra è condiviso da una larga maggioranza degli intervistati, i quali individuano nello Stato (61%) e nei suoi rappresentanti politici il principale soggetto cui ascrivere la responsabilità per eventuali uccisioni ingiustificate. Questo orientamento non contrasta con la filosofia del *DIU* che, responsabilizzando i soggetti operativi a ogni livello, intende prevenire l'equivoco invocato dagli imputati di Norimberga per sfuggire al peso delle proprie colpe. Qui a essere chiamata in causa è la responsabilità degli Stati e degli organi (governi, parlamenti) che li rappresentano, i quali sono tenuti a prevenire decisioni i cui effetti nocivi ricadrebbero su tutti. Da parte dei cittadini sale una domanda di trasparenza, che mal si concilia con scenari di guerra futura dominati da automatismi opachi e decisioni algoritmiche.

In secondo luogo, è evidente la diffidenza verso le promesse tecnologiche legate all'utilizzo di armi autonome. Solo una minoranza (32%) crede che l'impiego di questi sistemi possa contribuire a ridurre il numero di vittime civili nei conflitti, mentre la maggioranza relativa (61%) teme un effetto opposto, con un aumento delle vittime dovuto all'abbassamento della soglia di ricorso alla forza e alla perdita del controllo umano sulle decisioni critiche. Si tratta di un dato rilevante, che segnala la difficoltà delle narrazioni tecnocratiche nell'ottenere un consenso diffuso quando in gioco vi sono questioni di vita o di morte.

Proseguendo nell'analisi, le minacce più temute sono l'impiego degli *AWS* da parte di gruppi non statali, i malfunzionamenti tecnici, l'inefficacia del controllo umano e la maggiore facilità di escalation dei conflitti autonomizzati. A queste preoccupazioni si aggiunge la questione della governance: sebbene non vi sia un'istituzione internazionale identificata in modo univoco come l'attore legittimato a intervenire, la maggior parte dei cittadini (64%) si orienta verso soluzioni multilaterali e sovranazionali esprimendo fiducia, in ordine crescente, nell'Unione Europea, nei trattati internazionali e nelle Nazioni Unite.

La netta preferenza per il *Controllo Umano Significativo* – considerato necessario da quasi 2/3 del campione – rappresenta il fulcro di un'opposizione non tanto tecnica, quanto intrinsecamente etico-politica. In un'epoca in cui la tecnologia tende a permeare ogni ambito della vita, la possibilità che anche le decisioni letali vengano delegate a sistemi autonomi genera resistenze trasversali, le quali attraversano il tessuto sociale indipendentemente da età, genere e orientamento ideologico.

***

Come abbiamo visto in quest'ultima parte dedicata alla sfida sociale, l'allineamento dell'opinione pubblica del nostro Paese nei confronti di molte delle posizioni degli esperti espresse nelle parti precedenti – dedicate rispettivamente alla sfida etica, a quella tecnologica e a quella giuridico-diplomatica – è notevole e induce a riflettere.

Indubbiamente a questo allineamento ha contribuito la costruzione del questionario sulla base di una concomitante riflessione e scambio tra gli autori del Rapporto – i ricercatori delle discipline scientifico-tecnologiche e i ricercatori delle discipline sociali – iniziati già alcuni anni fa (Farruggia, 2023). L'allineamento tematico, però, non comportava automaticamente la convergenza dei contenuti che è stata invece riscontrata.

Una più accurata verifica delle origini di tale convergenza sarebbe senz'altro utile, ma nel frattempo è un dato acquisito che molte delle preoccupazioni degli esperti sono proprie anche dei profani. Data la complessità degli argomenti trattati e, contemporaneamente, data la scarsità di approfondimenti offerti al cittadino da chi potrebbe/dovrebbe fornirli (il ceto politico e i mass media tra i primi) è qualitativamente importante ma quantitativamente contenuta l'influenza *diretta* degli esperti. In generale, e in Italia non meno che in altri Paesi occidentali, gli allarmi sui rischi della ricerca e sviluppo della IA militare lanciati in più occasioni da autorevoli scienziati, destano un'eco scarso nei media tradizionali, in particolare giornali e programmi televisivi di orientamento mainstream. I punti di contatto con le competenze dei ricercatori e le loro preoccupazioni in riferimento alle sfide poste dalla IA militare sono quindi da spiegare con il *sentiment* dell'uomo della strada, istintivamente diffidente verso l'uso della forza in genere e quindi anche della strumentalizzazione politica e strategica dell'innovazione tecnologica in tale dominio.

Ripercorrendo le tappe delle analisi proposte nel Rapporto in una modalità quasi letterale (seppure semplificata), il campione da noi intervistato ha espresso nella sua maggioranza assoluta forti preoccupazioni circa le sfide poste dall'applicazione della IA alla guerra. Per questo 3/4 degli intervistati si dichiarano molto o abbastanza contrari alle armi letali autonome, nella misura del 60% si oppongono al suo impiego, più di 2/3 condividono la preoccupazione etica circa il controllo umano insufficiente e la conseguente incertezza delle responsabilità. È poi significativo che tra le misure per regolamentare questo tipo di armi, oltre al 17% che cita la sensibilizzazione dell'opinione pubblica, quasi 2/3 dei rispondenti privilegino le sedi e gli strumenti multilaterali resi disponibili dal sistema delle Nazioni Unite e dalla diplomazia.

In conclusione l'insieme della società civile in tutte le sue articolazioni – gli scienziati con le loro esternazioni, le ONG, le associazioni, le Chiese con le loro iniziative, i cittadini con l'espressione delle proprie opinioni – lanciano un no forte e chiaro all'accettazione passiva della logica della guerra autonomizzata. Al contrario, dalla società civile proviene una posizione articolata e motivata, che chiede trasparenza, controllo democratico e rispetto dei diritti fondamentali anche (e soprattutto) nei contesti di conflitto armato. I governi e le istituzioni internazionali hanno di fronte a sé un'ulteriore e decisiva sfida: tenere conto della voce della società, non solo come parametro di legittimità, ma anche come fattore attivo nel processo di regolazione delle nuove tecnologie belliche. In un

mondo sempre più attraversato da tensioni geopolitiche e rapide innovazioni, la tutela della dignità umana e del diritto deve restare il principio guida, anche di fronte alla seduzione dell'efficienza algoritmica.

Riferimenti bibliografici

Altmann, J. (2013). Arms control for armed uninhabited vehicles: an ethical issue. *Ethics and Information Technology* 15.2 (2013), 137-152.

Altmann, J., & Sauer, F. (2017). Autonomous weapon systems and strategic stability. *Survival,* 59(5), 117-142.

Arkin, R.C. (2009). *Governing lethal behavior in autonomous robots*. London: Chapman & Hall/CRC.

Asaro, P. (2012). On banning autonomous weapon systems: Human rights, automation, and the dehumanization of lethal decision-making. *International Review of the Red Cross*, 94(886), 687-709.

Battistelli, F. (2024). Dalle macchine di Turing alle macchine di Stranamore? I rischi nelle applicazioni della AI alla sicurezza internazionale e interna. In *Iriad Review* 01/2024, 4-16

Battistelli, F., & Farruggia, F. (2025). "Public opinion between rearmament and crisis of the nuclear taboo". In *AA.VV., Nuclear Weapon Risks in a Problematic Time, Springer, Berlin*. (in corso di stampa).

Bucchi, M., Pellegrini, G., Rubin, A., & Saracino, B. (a cura di). (2024). *Annuario Scienza tecnologia e Società*. Bologna: Il Mulino.

Bulletin of the Atomic Scientists. (2025). *Doomsday Clock set at 89 seconds to midnight*. Disponibile a: https://thebulletin.org/doomsday-clock/2025-statement/.

Caren, N., Ghoshal, R. A., & Ribas, V. (2011). A social movement generation: Cohort and period trends in protest attendance and petition signing. *American Sociological Review*, 76(1), 1-27.

Center for AI Safety. (2023). *Statement on AI Risk*. Disponibile a: https://safe.ai/work/statement-on-ai-risk.

CICR - Comitato Internazionale della Croce Rossa. (2024). *International humanitarian law and the challenges of contemporary armed conflicts*. International Committee of the Red Cross. Disponibile a: https://www.icrc.org/en/publication/international-humanitarian-law-and-challenges-contemporary-armed-conflicts-building.

Di Giuseppe, M., Paula, K., & Rommel, T. (2025). *AI on the Battlefield? Revisiting Public Support for LAWs*. Open Science Framework.

Farruggia, F. (a cura di). (2023). *Dai droni alle armi autonome. Lasciare l'Apocalisse alle macchine?* Milano: FrancoAngeli.

Fondazione Leonardo Civiltà delle Macchine & Centro Studi Americani. (2022). "Winning the Artificial Intelligence Era. Quantum Diplomacy and the Power of Automation". Presentato nell'ambito del convegno tenutosi il 27 aprile 2024 a Roma

presso il Centro Studi Americani. Disponibile a: https://www.fondazioneleonardo.com/sites/default/files/downloads/2024-05/Quantum_Diplomacy_ITA.pdf.

Future of Life Institute. (2015). *Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter*. Disponibile a: https://futureoflife.org/open-letter/ai-open-letter/.

G7 Science Academies. (2019). *Artificial Intelligence and Society*. Disponibile a: https://www.lincei.it/sites/default/files/attachments/2019-G7-Artificial_intelligence_and_society.pdf.

Galliott J., Baggiarini, B., & Rupka S. (2021). Empirical data on attitudes toward autonomous systems. In Galliott J., MacIntosh D., & Ohlin J. D. (eds.), *Lethal autonomous weapons: Re-examining the law and ethics of robotics warfare*. Oxford University Press, 137–158.

ICRAC - International Committee for Robot Arms Control. (2013). *Computing experts from 37 countries call for ban on killer robots.* Disponibile a: https://www.icrac.net/computing-experts-from-37-countries-call-for-ban-on-killer-robots/.

Ipsos. (2021). *Global attitudes towards autonomous weapons*. Disponibile a: https://www.ipsos.com.

Lushenko, P. (2023). *AI and the future of warfare: The troubling evidence from the US military*. Bulletin of the Atomic Scientists.

Parisi, G. (2023). Prefazione. In Farruggia, F. (a cura di). *Dai droni alle armi autonome. Lasciare l'Apocalisse alle macchine?* Milano: FrancoAngeli, 13-17.

PAX. (2018). *Belgium votes to ban killer robots*. Disponibile a: https://paxforpeace.nl/news/overview/belgium-votes-to-ban-killer-robots.

Pekarev, J. (2023). Attitudes of military personnel towards unmanned ground vehicles (UGV): a study of in-depth interview. *Discover Artificial Intelligence*, 3(24), https://doi.org/10.1007/s44163-023-00058-4.

Qiao-Franco, G., & Bode, I. (2023). Weaponised Artificial Intelligence and Chinese Practices of Human–Machine Interaction. *The Chinese Journal of International Politics*, *16*(1), 106–128.

Rosendorf, O., Smetana, M., & Vranka, M. (2022). Autonomous weapons and ethical judgments: Experimental evidence on attitudes toward the military use of "killer robots". *Peace and Conflict: Journal of Peace Psychology*, *28*(2), 177–183.

Rosendorf, O., Smetana, M., Vranka, M. & Dahlmann, A. (2024). Mind over metal: Public opinion on autonomous weapons in the US, Brazil, Germany, and China. *SSRN Working Paper*.

RTBF. (2017). *Dans une lettre ouverte, des scientifiques demandent l'interdiction des robots tueurs*. Disponibile a: https://www.rtbf.be/article/dans-une-lettre-ouverte-des-scientifiques-demandent-l-interdiction-des-robots-tueurs-9781662.

Scharre, P. (2018). Army of none: Autonomous weapons and the future of war. *W.W. Norton & Company.*

SonntagsBlick. (2021). *Swiss researchers warn about autonomous weapons*. Disponibile a: https://www.swissinfo.ch/eng/business/swiss-researchers-warn-about-autonomous-weapons/48379426.

Sparrow, R. (2007). Killer robots. *Journal of Applied Philosophy*, 24(1), 62-77.

Taddeo, M. & Blanchard, A. (2022). A Comparative Analysis of the Definitions of Autonomous Weapons Systems. *Science and Engineering Ethics*, 28 (5), 1-22.

Tamburrini, G. (2020). *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*. Roma: Carocci

Tamburrini, G. (2023). Governing lethal autonomy: The challenge of meaningful human control. *AI and Ethics*, 3(2), 145-158.

The Brussels Times. (2023). *Belgium upholds decision to ban killer robots*. Disponibile a: https://www.brusselstimes.com/350980/belgium-upholds-decision-to-ban-killer-robots.

The Guardian. (2017). *Ban killer robots, experts urge Australian and Canadian leaders*. Disponibile a: https://www.theguardian.com/science/2017/nov/06/ban-killer-robots-experts-urge-australian-and-canadian-leaders-malcolm-turnbull-justin-trudeau.

USPID - Unione Italiana Scienziati per il Disarmo. (2019). *Gli scienziati italiani: fermate la corsa ai robot killer*. Disponibile a: https://retepacedisarmo.org/stop-killer-robots/2019/03/gli-scienziati-italiani-fermate-la-corsa-ai-robot-killer/.

Zhang, B., Anderljung, M., Kahn, L., Dreksler, N., Horowitz, M. C., & Dafoe, A. (2021). Ethics and governance of artificial intelligence: Evidence from a survey of machine learning researchers. *Journal of Artificial Intelligence Research*, 71, 591-666.

Istituto di Ricerche Internazionali ARCHIVIO DISARMO
Via Paolo Mercuri, 8
00193 Roma
info@archiviodisarmo.it