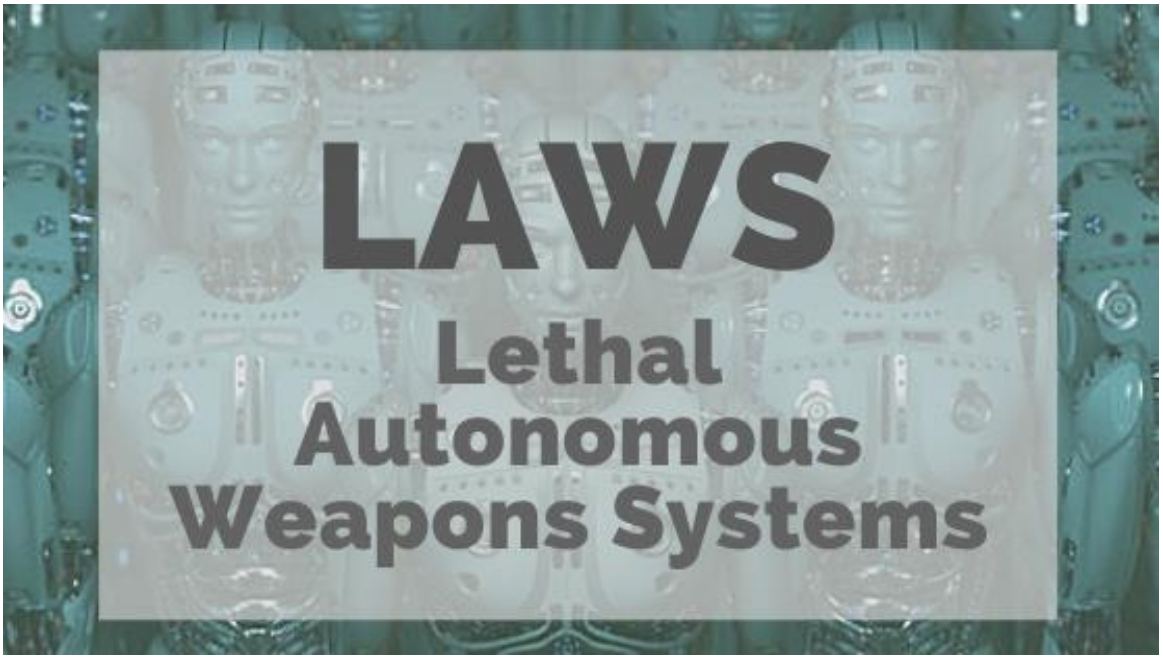
IRIAD REVIEW
Studi sulla pace e sui conflitti

Periodico mensile IRIAD - ISSN 2611-3953 n. 7-8 luglio/agosto 2020

La questione delle armi letali autonome e le possibili azioni italiane ed europee per un accordo internazionale



INDICE

Abstract p. 5
Sintesi p. 8

Introduzione
Miti, sogni e incubi dell'autonomia delle armi p. 11
Bibliografia p. 18

Cap. 1. *Aspetti filosofici, etici e sociali delle armi autonome* p. 20
1.1 Le fonti normative del dibattito etico sulle armi autonome p. 20
1.2 Autonomia operativa e principi di distinzione e proporzionalità p. 21
1.3 Armi autonome, responsabilità e dignità degli esseri umani p. 25
1.4 Armi autonome ed etica delle conseguenze p. 26
Bibliografia p. 31

Cap. 2. *Caratteristiche, prospettive e problematicità dell'Intelligenza Artificiale* p. 33
2.1 Introduzione p. 33
2.2 Machine Learning p. 35
2.3 Limiti e problematicità del Machine Learning p. 40
2.4 Conclusioni p. 46
Bibliografia p. 47

Cap. 3. *Applicazioni militari e i potenziali sviluppi dei LAWS* p. 50
3.1 Introduzione p. 50
3.2 Stato dell'arte p. 56
3.3 Munizioni Loitering p. 56
3.4 I sistemi d'arma ravvicinata o a corto raggio (CIWS) p. 60
3.5 Aerei da combattimento senza equipaggio UAV p. 63
3.6 Munizioni guidate di precisione p. 66
3.7 Veicoli terrestri senza equipaggio - UGV p. 70
3.8 Veicoli Marini senza equipaggio - UMV p. 74
3.9 Controllo delle frontiere p. 76
3.10 SWARM p. 79
3.11 Conclusioni p. 82

Acronimi p. 83
Bibliografia p. 84

Cap. 4. *Vulnerabilità delle tecnologie informatiche e AWS* p. 87
4.1 Introduzione p. 87
4.2 Vulnerabilità intrinseche delle tecnologie informatiche p. 88
4.3 Vulnerabilità specifiche delle tecnologie IA/ML p. 91
4.4 Dibattito Internazionale p. 95
4.5 Conclusioni p. 97
Bibliografia p. 99

Cap. 5. *L'opinione pubblica e le armi autonome* p. 101
5.1 Opinione pubblica e uso della forza p. 101
5.2 L'opinione pubblica mondiale di fronte alle armi autonome p. 102
5.3. L'opinione pubblica italiana e le armi autonome p. 104
5.3.1 L'indagine IPSOS p. 104
5.3.2 L'indagine di Archivio Disarmo p. 105
5.4 Le giovani generazioni e le armi autonome p. 110
5.5 Osservazioni conclusive p. 113
Bibliografia p. 114

Cap. 6
Il diritto internazionale umanitario e la sfida delle armi autonome all'intus-legere p. 115
6.1 Introduzione p. 115
6.2 Mezzi e metodi di combattimento p. 116
6.3 La protezione dei civili p. 118
6.4 Il ruolo dei consulenti giuridici e politici p. 120
6.5 L'umanità e la pubblica coscienza p. 121
6.6 La posizione UE e italiana p. 123
6.7 Conclusioni p. 127
Bibliografia p. 129

Cap. 7
Il dibattito in seno alla Convenzione su certe armi convenzionali CCW a Ginevra p. 131
7.1 La comunità scientifica e i problemi etici sollevati dalle armi autonome p. 131
7.2 Le armi autonome alla *Convention on Certain Conventional Weapons* p. 132
7.3 Il controllo umano significativo e i suoi contenuti p. 135
7.4 Verso un trattato internazionale sul controllo umano significativo delle armi autonome? p. 136
Bibliografia p. 141

Cap. 8
Il ruolo della società civile p. 143
8.1 Le iniziative della società civile p. 143
8.2 Conclusioni p. 146
Bibliografia p. 147

Cap. 9
Considerazioni finali. Linee d'azione e di controllo a livello nazionale e europeo: possibilità e potenzialità p. 149
9.1 La Rome Call for AI Ethics p. 149
9.2 L'IA nel settore degli armamenti p. 149
9.3 Criticità e rischi dell'IA nel settore bellico p. 150
9.4 Gli aspetti giuridici p. 152
9.5 Il dibattito e le proposte a livello internazionale p. 153
9.6 L'azione italiana in sede multilaterale p. 153
Bibliografia p. 157

Gli autori p. 158

Il gruppo di ricerca, coordinato da Maurizio Simoncelli (cap. 9), è costituito da Fabrizio Battistelli (introduzione), Adriano Iaria e Sofia Bertieri (cap. 6), Francesca Farruggia (cap. 5), Barbara Gallo (cap. 8), Juan Carlos Rossi (cap. 3), Gian Piero Siroli (cap. 2, 4) e Guglielmo Tamburrini (cap. 1,2 e 7). La ricerca dell'Istituto di Ricerche Internazionali Archivio Disarmo IRIAD si è avvalsa della collaborazione dell'Unione degli Scienziati Per Il Disarmo USPID e di suoi membri, il prof. Gian Piero Siroli (Univ. di Bologna, CERN, USPID) e il prof. Guglielmo Tamburrini (Univ. di Napoli Federico II, ICRAC, USPID). Gli autori sono profondamente grati al dott. Diego Latella dell'Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" del CNR e membro dell'USPID, per l'intenso confronto su varie problematiche qui affrontate, per la generosa condivisione di una vasta documentazione e per i suoi preziosi suggerimenti e commenti sulle versioni preliminari di questo rapporto.
Il presente report è stato redatto con il sostegno finanziario dell'Unità di Analisi, Programmazione, Statistica e Documentazione Storica del Ministero degli Affari Esteri e della Cooperazione Internazionale, ai sensi dell'art.23bis del DPR 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, né quelle dell'USPID.

IRIAD REVIEW. Studi sulla pace e sui conflitti. - ISSN 2611-3953

Mensile dell'IRIAD (Istituto di Ricerche Internazionali Archivio Disarmo)

Via Paolo Mercuri 8, 00193 – Roma (RM)

C.F. 97018990586, P.Iva 04365231002 Tel. + 39 06 36000343
info@archiviodisarmo.it - www.archiviodisarmo.it

Direttore Editoriale: Maurizio Simoncelli

Direttore Responsabile: Fabrizio Battistelli

Registrazione Tribunale di Roma n. 53/2018

Copyright © IRIAD (Istituto di Ricerche Internazionali Archivio Disarmo)

The question of Lethal Autonomous Weapons Systems- LAWS and possible Italian and European action to obtain international agreement

Abstract

This report is the work of a research group formed by researchers from two non-profit organisations– IRIAD-Institute of international research ARCHIVIO DISARMO and USPID-Scientists' Union for Disarmament – active in Italy for almost forty years in an analysis of and advocacy for peace terms and arms reduction and control. Research carried out for the Ministry of Foreign Affairs and International Co-operation*, has confronted the question of Artificial Intelligence as applied to armaments (LAWS)**.** After dealing with the historic, culturally-symbolic, philosophic and ethical aspects concerning the possibility of entrusting, all or in part, fighting activity to entities other than humans, the report examines some characteristics of Artificial Intelligence and their military application. These become evident in the research, development, operational ability and adoption of autonomous lethal weapons by the armed forces and the major powers (and possibly not only by them) in the foreseeable future. At the same time possible alternatives to an impending proliferation and race for these new types of armament that can be used and are workable in two different, but equally important, ambits have been examined. That of civil society is represented by the attitude of public opinion and by the position taken by the international scientific community. That of governments is represented by diplomatic talks (currently held in Geneva by the Group of Governmental Experts on LAWS) aimed an agreement for the adoption of measures for the prevention and control of autonomous lethal weapons.

Artificial Intelligence (AI) is transforming, if not revolutionising, our society ever more, offering multiple and important opportunities in all sectors. AI is also experiencing a widespread and growing application in the military sector. In this latter area its application to weapon systems, although relatively initial, is already a reality in military R&D in numerous states from the USA to Russia, from France to Great Britain and from South Korea to Israel.

Various weapon systems are undergoing ever more advanced applications, stretched to create war machines capable not only of operating in an ever growing autonomous way, but also in a rather shorter time span than the human one. Although many of these weapon systems are still in a research and development stage or in prototype form, others are already operative, above all in the defence sector, such as the Russian S-400 (since 2007) or the United States MIM 404 Patriot.

For some years in this sector, interested governments have spent a considerable amount of resources, that appear to be growing exponentially, so much so that the predictions for the impact of AI (civil and military) are about $13 trillion by 2030.

At present, AI still shows substantial limitations. Different experimentation, using

the adversarial testing method, have highlighted how the neural network can be led into error, mistaking a school bus and a dog for an oyster or a tortoise for a rifle. In an analogous way the analysis of the process of machine learning (ML) has brought out some critical points that cannot be neglected, to the point that it appears risky to trust such technologies as a single decisional tool, at least as the present level of technology stands.

Among other things, it has emerged that AI and ML systems are vulnerable both to attacks of an exploratory, causative and catastrophic-forgetting nature.

Elevated levels of risk emerge connected both to an adversarial offensive attack that can in various ways induce errors in the system, and to an internal malfunction within AI that is difficult to foresee due to its own updating processes and to the complex functioning of the algorithms.

The interest aroused by LAWS among various governments and major powers is justified by a multiplicity of motives (greater precision in attacks with respect to human ones, elevated reactive speed to an attack suffered, absence of emotive human stress, reduction in collateral damage with regards to international human rights, tighter rules of engagement etc.). On the other hand, one can counter that the lower loss of the life for one's own military could create an incentive both for an acquiescence of public opinion and for that of the political decision makers towards new armed conflict, founded on the conviction that war need only be fought by machines. Among other things, the major industrial (and military) powers could be induced to believe themselves easy victors in wars against countries not possessing LAWS.

Furthermore, given the relatively reduced costs of robotics and of AI, not only could a new arms race develop involving an increased number of states, but also the spread of such weapon systems to irregular military formations and to terrorist organisations with unpredictable consequences for internal and external security. If we then also consider progress in the 3D printer sector and the dangers connected to the evolution of technology and materials, the risks appear still more imminent.

Certainly the use of autonomous weapon systems would bring with them a radical transformation of conflict management, with the inevitable subversion of the picture in ethical, juridical, behavioural and operational terms. On the other hand, civil society has for some time warned of the risks and dangers of an uncontrolled application of such technology to weapon systems, as became apparent from the opinion polls of 2018 and 2019. Ever since 2012 the scientific community has, on many occasions, taken a very strong position against it.

From a juridical point of view, LAWS must however refer both to compliance with the norms of international human rights, to the maintenance of the chain of responsibility for warlike action, and to ethical questions represented by a respect for the dignity of those human beings involved in the conflict. In war, and especially one using LAWS, it is rather difficult to apply not only the principle of distinction, but also of

that of proportionality. Even more, in the case of a mistake with LAWS, it seems difficult to trace the relative responsibility.

On the other hand, also governments are confronting each other in a multilateral ambit, in the talks on CCW (Convention on Certain Conventional Weapons) at Geneva, there are rather different positions among those who are absolutely against, among those who are in favour, and among those who seek a concrete operative mediation.

From the international debate taking place at the CCW, the concept of "Meaningful Human Control - MHC) seems to be emerging, which appears to find a certain consensus among different governmental delegations, because it also avoids giving up hope in the debate on what LAWS may be and concentrating instead on controlling it and on its different levels.

Italy has not yet produced any official document on the matter, but in 2019 Defence Minister Mr. Lorenzo Guerini declared that the “perplexity over the risky use of artificial intelligence devices". While the European Parliament approved a resolution on 27 February 2014 totally against autonomous weapons, in which they requested banning "the development, production and use of fully autonomous weapons which enable strikes to be carried out without human intervention", the EU Council of November 2016 requested a regulation inspired by international law and in particular by international human rights.

Also within the ambit of NATO, in the report of the Science And Technology Committee (STC) - Sub-Committee on Technology Trends and Security (STCTTS) of 2019, the importance is underlined, from now on, of tackling ethical, legal and social questions from the standpoint of an adequate human involvement in decisions involving the use of force, without leaving such decisions only to the allied Armed forces**.**

For Italy, on the other hand, important room for action presents itself in tune with what has emerged from various meetings. The Italian government, in such a sense, can play an important mediation and connecting role between the different positions held by different countries. It can raise awareness of the critical nature of such technology and its consequent risks, with the purpose of reaching an international agreement capable of avoiding a danger deriving from anarchy in a proliferation of these new weapon systems.

* The present report has been drawn up with the financial backing of the Ministry of Foreign Affairs and International Co-operation under Article 23b of DPR 18/1967. The data and statements contained in this report are the responsibility of the authors who form the research group and do not involve the Minister of Foreign Affairs and International Co-operation of the Italian Republic.

La questione delle armi letali autonome e le possibili azioni italiane ed europee per un accordo internazionale

Sintesi

Il presente rapporto è opera di un gruppo di ricerca formato da ricercatori delle due organizzazioni non profit – IRIAD-Istituto di ricerche internazionali ARCHIVIO DISARMO e USPID-Unione Scienziati per il Disarmo – attive da quasi un quarantennio nell'analisi e nell'advocay sui temi della pace e del controllo e riduzione degli armamenti (arms reduction and control). La ricerca, effettuata per conto del ministero degli Affari esteri e della cooperazione internazionale*, ha affrontato la questione dell'Intelligenza Artificiale applicata agli armamenti (LAWS). Dopo una trattazione degli aspetti storici, simbolico-culturali, filosofici ed etici relativi alla possibilità di affidare in tutto o in parte le attività di combattimento a entità diverse dagli uomini, il Rapporto esamina alcune caratteristiche dell'Intelligenza Artificiale e delle sue applicazioni militari quali prendono corpo nella ricerca, sviluppo, operatività e adozione delle armi letali autonome da parte delle forze armate delle principali potenze (ed eventualmente non soltanto da esse) oggi e nel prevedibile futuro. Contestualmente vengono esaminate le possibili alternative a una incombente proliferazione e corsa a tali armamenti di nuovo tipo, praticate e praticabili in due diversi, e ugualmente importanti, ambiti. Quello della società civile è rappresentato dagli atteggiamenti dell'opinione pubblica e dalle prese di posizione della comunità scientifica internazionale; quello dei governi, dai colloqui diplomatici finalizzati ad accordi per l'adozione di misure di prevenzione e controllo nei confronti delle armi letali autonome.

L'Intelligenza Artificiale (IA) sta trasformando, se non rivoluzionando, sempre più la nostra società offrendo molteplici e rilevanti opportunità in tutti i settori. L'IA è in via di larga e crescente applicazione anche nel settore militare. In quest'ultimo campo, le sue applicazioni ai sistemi d'arma, seppur relativamente agli esordi, sono già una realtà in ambito industriale e militare presso numerosi stati, dagli USA alla Russia, dalla Francia alla Gran Bretagna, dalla Corea del Sud ad Israele.

Svariati sistemi d'arma sono sottoposti ad applicazioni sempre più avanzate, tese a realizzare macchine belliche capaci non solo di operare in una prospettiva di crescente autonomia, ma anche in uno spazio temporale assai più breve di quello umano. Se molti di questi sistemi d'arma sono ancora in via di ricerca e sviluppo o sotto forma di prototipi, altri sono già operativi soprattutto nel settore difensivo, come l'S-400 russo (dal 2007) o il MIM 404 Patriot statunitense.

Da alcuni anni ingenti risorse vengono destinate dai governi interessati in questo settore, che appare destinato ad una crescita esponenziale, tanto che le previsioni per l'impatto economico dell'IA (civile e militare) si aggirano intorno ai $ 13 trilioni entro il 2030.

Allo stato attuale l'IA presenta ancora forti limiti. Diverse sperimentazioni, utilizzando la metodologia di *adversarial testing,* hanno messo in rilievo come la rete neurale possa essere indotta in errore, scambiando uno scuolabus e un cane per un'ostrica o una tartaruga per un fucile. Analogamente l'analisi del processo di *machine learning* (ML) ha fatto emergere alcuni punti critici che non possono essere trascurati, tanto che risulta assai rischioso affidarsi a tali tecniche come *unico strumento* decisionale, stando almeno ai livelli tecnologici attuali.

Sono, tra l'altro, emerse vulnerabilità dei sistemi di IA e ML connesse sia a*exploratory attacks*, sia a*causative attacks*, sia a *catastrophic forgetting*. Emergono elevati rischi connessi sia ad un'azione offensiva avversaria che con varie modalità può indurre in errore il sistema, sia a malfunzionamenti interni e difficilmente prevedibili dell'IA dovuti ai propri processi di autoapprendimento e al complesso funzionamento degli algoritmi.

L'interesse destato dai LAWS presso vari governi e stati maggiori è giustificato con molteplici motivi (maggiore precisione negli attacchi rispetto a quella umana, elevata rapidità reattiva ad un attacco subito, assenza di stress emotivo umano, riduzione dei danni collaterali nel rispetto del diritto umanitario internazionale, regole d'ingaggio più stringenti ecc.). D'altro canto, si può opporre che le minori perdite di vite dei propri militari potrebbero incentivare sia l'acquiescenza dell'opinione pubblica, sia quella dei decisori politici verso nuovi conflitti armati, fondandosi sulla convinzione di una guerra combattuta solo dalle macchine. Tra l'altro, le maggiori potenze industriali (e militari) potrebbero essere indotte a ritenersi facili vincitrici di guerre contro Paesi non dotati di LAWS.

Inoltre, dati i costi relativamente ridotti della robotica e dell'IA, potrebbero svilupparsi non solo una nuova corsa al riarmo coinvolgente un numero elevato di stati, ma anche una diffusione di tali sistemi d'arma presso milizie irregolari e formazioni terroristiche, con conseguenze imprevedibili sulla sicurezza interna ed esterna. Se poi consideriamo anche i progressi nel settore delle stampanti 3D e i pericoli connessi all'evoluzione tecnologica e dei materiali, i rischi appaiono ancora più incombenti.

Certamente l'uso di sistemi d'arma autonomi comporterebbe una trasformazione radicale nella gestione dei conflitti, con l'ineludibile sovvertimento del quadro in ambito etico, giuridico, comportamentale e operativo. Peraltro, la società civile da tempo ha avvertito i rischi e i pericoli di un'applicazione incontrollata di tale tecnologia ai sistemi d'arma, come è emerso da indagini demoscopiche del 2018 e del 2019. A partire dal 2012 la comunità scientifica ha, più volte e in misura massiccia, preso posizione contraria.

Dal punto di vista giuridico, i LAWS devono comunque fare riferimento sia all'ottemperanza delle norme del diritto internazionale umanitario, sia al mantenimento della catena delle responsabilità nelle azioni belliche, sia alla questione etica rappresentata dal rispetto della dignità degli esseri umani coinvolti nel conflitto.

In guerra, e specialmente ad opera di un LAWS, è assai difficile applicare non solo il principio di distinzione, ma anche quello di proporzionalità. Ancor di più, in caso di errore del LAWS, appare difficile risalire alle relative responsabilità.

Pertanto, anche i governi si stanno confrontando in ambito multilaterale, nei colloqui sulla CCW (*Convention on Certain Conventional Weapons*) a Ginevra, dove vi sono posizioni assai differenziate tra chi è assolutamente contrario, tra chi è decisamente favorevole e tra chi punta ad una mediazione operativamente concreta.

Dal dibattito internazionale presso la CCW sembra emergere il concetto del "controllo umano significativo" (Meaningful Human Control - MHC), che appare trovare un certo consenso presso diverse delegazioni governative, anche perché evita di disperdersi nel dibattito su cosa sia un LAWS, concentrandosi invece sul suo controllo e sui suoi diversi livelli.

L'Italia non ha ancora prodotto nessun documento ufficiale in merito, ma nel 2019 il ministro della difesa Guerini ha dichiarato le "perplessità sull'utilizzo spinto dei dispositivi di intelligenza artificiale". Mentre il Parlamento Europeo nel 2014 ha approvato una risoluzione nettamente contraria alle armi autonome, in cui veniva richiesto di "vietare lo sviluppo, la produzione e l'impiego di armi completamente autonome che consentono di sferrare attacchi senza alcun intervento umano", il Consiglio Ue del novembre 2016 invece ha richiesto una regolamentazione ispirata dal diritto internazionale e in particolare dal diritto internazionale umanitario.

Anche in ambito NATO, nel rapporto allo Science And Technology Committee (STC) - Sub-Committee on Technology Trends and Security (STCTTS) del 2019, si sottolinea l'importanza di affrontare tutte le questioni etiche, legali e sociali sin d'ora, nella prospettiva di un adeguato coinvolgimento umano nelle decisioni sull'uso della forza, senza lasciare tali decisioni solo alle Forze armate alleate.

Per l'Italia si presentano pertanto degli importanti spazi d'azione in sintonia con quanto è emerso dai vari consessi. Il governo italiano, in tal senso, può svolgere un ruolo importante di mediazione e di raccordo tra le diverse posizioni degli Stati evidenziando le criticità di tale tecnologia e i rischi conseguenti, al fine di giungere ad un accordo internazionale condiviso in grado di evitare una pericolosa deriva anarchica nella proliferazione di questi nuovi sistemi d'arma.

* Il presente report è stato redatto con il sostegno finanziario del Ministero degli Affari Esteri e della Cooperazione Internazionale, ai sensi dell'art.23bis del DPR 18/1967. I dati e le affermazioni contenute nel presente rapporto sono responsabilità degli autori che hanno formato il gruppo di ricerca e non coinvolgono il Ministero degli Affari Esteri e della Cooperazione internazionale della Repubblica italiana.

Introduzione
Miti, sogni e incubi dell'autonomia delle armi

Affrontando la questione delle *LAWs – Lethal Authonomous Weapons*, il sogno degli eserciti (occasionale nei soldati, ricorrente nei comandanti) è stato quello di un *Deus ex machina* (qualcuno o qualcosa) che intervenisse nella battaglia gettandovi tutto il suo peso e inclinando il piatto della bilancia a favore del nostro esercito e a scapito di quello nemico.

Quel trattato di polemologia in versi che è l'Iliade offre un esempio ineguagliato di questa aspettativa umana. Senza risparmio di energie umane e divine, una panoplia di interventi prodigiosi fatti di apparizioni, azioni di forza, trucchi e incantesimi vede gli dei dell'Olimpo schierarsi a favore dei greci (Era, Atena, Poseidone) e, sul fronte opposto, dei troiani (Apollo, Afrodite). Sul piano storico, all'incirca cinque secoli più tardi i Romani vittoriosi inizieranno a trasportare nell'Urbe non soltanto i re e i capi dei popoli sconfitti per farli comparire in catene al seguito del generale trionfatore, ma anche i loro dèi con i relativi poteri, da includere direttamente nel Pantheon romano. Agli albori dell'età moderna, il ruolo che nell'antichità classica era stato degli dèi verrà sempre più spesso rivestito dalle forze della natura. Sono numerosi gli esempi di eventi naturali apparsi a soccorrere l'esercito degli uni e a disperdere quello degli altri (l'esempio più noto è rappresentato dal mito dei marosi che proteggono Britannia e affondano i vascelli della *Invencible Armada* spagnola).

L'interrogativo è, giunti all'età dell'Illuminismo, come dare spazio al mito, che (insieme al suo stretto parente l'inganno) è un ingrediente della propaganda di guerra, utile a esaltare la volontà di vittoria della popolazione e dell'esercito, contemporaneamente deprimendo la volontà del nemico. Coniuga mito e pragmatismo il maggiore stratega del XVIII secolo, Federico II di Prussia, che punta su fattori organizzativi moderni quali l'addestramento e la disciplina dei reparti, con lo scopo di conseguire la superiorità sul campo di battaglia. È sintomatico che, con questo obiettivo operativo, il re prussiano evochi con il fedele generale von Saldern una metafora essa stessa molto moderna: quella di soldati da addestrare tanto e così bene al maneggio delle armi da trasformarli in "automi tiratori"[1].

Tuttavia si tratta di semplici premonizioni. Bisognerà aspettare la seconda metà del XX secolo perché, preparata dalle ingegnose applicazioni della scienza e della tecnologia a quella che Marx chiamava "l'arte di macellare gli uomini", faccia irruzione sulla scena bellica il concetto di automazione del campo di battaglia. Un cinquantennio fa, il 16 ottobre 1969, difronte al senato degli Stati Uniti il capo di stato maggiore dell'esercito, generale William C. Westmoreland disegnava il seguente scenario: "Nel

[1]In tema di automi è da ricordare che risalgono al Settecento non soltanto mode e trucchi ingannevoli come l'automa giocatore di scacchi detto "il Turco", ma anche i primi studi scientifici sull'automazione.

campo di battaglia del futuro le forze nemiche saranno identificate, tracciate e inquadrate come bersagli quasi istantaneamente mediante l'uso di connessione dati, valutazioni di intelligence assistite da computer e controllo di tiro automatizzato" (Westmoreland, 1969). Il successivo passo dottrinale in direzione del campo di battaglia automatizzato verrà compiuto un quarto di secolo più tardi con la dottrina della *Network-centricwarfare* (Abrahamsson e Ydén, 2005), mentre quello ancora successivo prende corpo oggi con le LAWs.

Come mai il generale Westmoreland esternava sotto forma di una solenne previsione quello che al momento non era altro che un *wishful thinking*? Non è un caso che l'urgenza di rendere automatico (e in prospettiva autonomo) il campo di battaglia sia stata formulata con tanta forza dal comandante del contingente di quella che, in proporzione alle forze impiegate, è stata la più sanguinosa e contestata guerra esterna nella storia degli Stati Uniti, cioè il Vietnam. Con 58mila soldati americani morti e oltre 300mila feriti Westmoreland ha avuto modo di osservare da vicino e toccare drammaticamente con mano i costi politici della carneficina.

A questo punto è necessario aprire una riflessione sulla più cruciale delle tendenze rintracciabili nel pensiero politico-strategico contemporaneo, in particolare presso la superpotenza mondiale Stati Uniti: il mito della guerra a perdite zero. L'idea di risparmiare il più possibile le vite dei propri soldati non è una novità per nessun esercito, naturalmente. È tuttavia nel XX secolo che l'ipotesi di realizzare un drastico risparmio di vite umane nei propri ranghi, massimizzando i risultati dell'azione militare (contemporaneamente massimizzando le perdite nei ranghi del nemico) si concretizza con l'avvento dell'arma aerea. Superata la fase del primo conflitto mondiale, in cui l'ambiente terrestre è dominato dai massacri simmetrici della guerra di trincea e quello aereo dai duelli cavallereschi tra i caccia a elica, il dopoguerra si caratterizza per l'invenzione da parte dell'italiano Giulio Douhet (1921), e nel 1937 per l'applicazione da parte dell'aviazione tedesca contro la città spagnola di Guernica, della dottrina del bombardamento “strategico”. Prendendo di mira non solo i bersagli militari, ma anche e soprattutto quelli civili, la teoria e la pratica del bombardamento strategico individuano nella moltiplicazione delle vittime tra la popolazione inerme non un semplice danno collaterale, ma l'obiettivo globale di demoralizzare il nemico e spezzare la sua resistenza. Nella seconda guerra mondiale bombardamenti a tappeto contro i centri abitati e altri obiettivi civili vengono inizialmente effettuati dai tedeschi a scopo terroristico contro Varsavia, Rotterdam, Coventry, Londra ecc.; non diversamente da come nella seconda metà del conflitto faranno gli Alleati per motivi tattici a Cassino e a scopo di ritorsione contro Colonia, Brema, Lubecca e Dresda, sino a toccare l'acme con l'irruzione sulla scena bellica dell'arma atomica. A fronte dell'accanita resistenza del territorio nazionale da parte delle forze armate giapponesi, le due bombe atomiche sganciate dagli Stati Uniti sulle città di Hiroshima e Nagasaki verranno ufficialmente giustificate con l'esigenza di risparmiare ulteriori perdite nelle file dei soldati americani.

Ancora tollerati in un contesto di emergenza estrema e globale come il secondo conflitto mondiale, gli enormi costi in vite umane presentati dalla guerra appaiono sempre meno legittimati agli occhi dell'opinione pubblica nei conflitti limitati della guerra fredda, quali la guerra di Corea (1954) e soprattutto il Vietnam (1964-1974). Per quanti sforzi faccia l'élite politica e militare per "vendere" ai cittadini elettori il sacrificio di vite umane, quest'ultimo appare ogni volta più spropositato in una società come quella contemporanea, progressivamente più secolarizzata e focalizzata sui diritti individuali. Le fasi storiche precedenti erano dominate dalla trascendenza religiosa che minimizzava la rilevanza della vita terrena in favore dell'unica vita veramente importante, quella ultraterrena. Invece, nella società contemporanea l'immanenza indotta dai valori individualistici della modernità e del mercato rende la morte intollerabile, un'entità "intrinsecamente vergognosa e orripilante", addirittura pornografica (Gorer, 1965, p. 171, cit. in Bauman, 1995). Un vero e proprio tabù, dunque difficilissimo da gestire e rischioso anche soltanto da nominare.

Al tempo stesso, in occidente i mass media esercitano nei confronti dei decisori politici uno "scrutinio" sempre più stringente (Burk, 1998), che ha raggiunto il suo apice proprio con la guerra del Vietnam. Documentando vividamente e in tempo pressoché reale ciò che accade nella giungla vietnamita (innanzitutto il sacrificio di migliaia di giovani coscritti che muoiono), a partire dall'offensiva del Tetscatenata dai Vietcong nel 1968, la tv americana porta "nei tinelli di casa" l'evidenza di una guerra che non avrebbe potuto essere vinta se non a un prezzo insostenibile per qualsiasi leader di una democrazia rappresentativa. Negli USA il crollo del consenso sarà devastante e determinerà il cedimento del fronte interno (polemiche politiche e mediatiche, opposizione di massa alla leva militare, oceaniche manifestazioni contro la guerra con in prima fila veterani e invalidi che restituiscono le decorazioni ecc.). Per un intero paese il trauma sarà profondo e la crisi per i vertici politici e militari degli Stati Uniti sarà inevitabile. Lo scandalo di uno spionaggio ai danni del partito di opposizione travolgerà lo stesso presidente Nixon e, dato specificamente rilevante, nelle politiche della difesa inizierà un lungo e tormentato processo autocritico che condurrà all'abbandono della coscrizione obbligatoria, all'introduzione di forze armate di soli volontari e a una drastica revisione delle capacità e metodologie comunicative delle forze armate[2].

Per quanto spregiudicata sia la loro azione di governo, in un ordinamento democratico rappresentativo gli esponenti delle élite possono sì cercare di condizionare in tutti i modi gli atteggiamenti dell'opinione pubblica (un caso di scuola resta la martellante campagna orchestrata nel 2002-03 dal presidente Bush per

[2] Sulla rivoluzione in materia di "gestione della notizia" (*news management*) attuata dai militari in occasione di interventi armati all'estero, si vedano le due campagne contro l'Iraq, quella del 1991 e quella del 2003.

legittimare l'intervento militare in Iraq con la giustificazione di dover neutralizzare le armi di distruzione di massa che sarebbero state in possesso di Saddam), ma non possono ignorare completamente la variabile del consenso. Un fattore decisivo nella concessione ovvero nel ritiro da parte del pubblico del sostegno nei confronti di un'azione militare è costituito dall'entità delle vittime – feriti e soprattutto deceduti – che essa è in grado di causare (Battistelli et al. 2012). A far riflettere i politici si adoperano gli studiosi che, non solo progressisti ma anche conservatori, appaiono unanimi nel diagnosticare una società occidentale che, nella sua fase postmoderna, è sempre meno incline all'uso della forza, tanto più quando provoca vittime. Fra gli stessi consulenti di George W. Bush, Robert Kagan (2004, p. 17) ammette l'esistenza di una "crescente avversione postmoderna alla forza militare, alla politica di potenza e all'idea stessa di equilibrio tra potenze". Ovvero, come precisa il clintoniano Joseph S. Nye (2005, p. 24), fautore del *soft power* USA basato su mass media, consumi e diritti umani, "le democrazie post-industriali sono focalizzate sul benessere piuttosto che sulla gloria delle gesta militari e aborriscono la perdita di vite umane".

Sociologi, politologi ed esperti di strategia hanno discusso a lungo tanto sulla genesi di questo che è ormai un consolidato atteggiamento occidentale di fronte alla morte e si sono interrogati sulle possibili "soluzioni", siano esse rappresentate dagli adattamenti spontanei degli attori interessati (i soldati e le loro comunità di riferimento) ovvero quelli intenzionali (cioè le strategie comunicative e organizzative) dei vertici politici e militari.

Per quanto riguarda le spiegazioni, è da ricordare quella "demografica" di Edward Luttwak (1994, p. 27), secondo cui paesi come quelli appartenenti al Consiglio di sicurezza dell'ONU (i tre occidentali e la Russia, ad esclusione della Cina, non citata) possono ancora possedere base economica e forza militare adeguate per condurre guerre, ma "le loro società sono così allergiche alle perdite, che di fatto essi sono debellicizzati o quasi". Operata questa diagnosi, l'autore descrive l'imperante atteggiamento di ipertutela delle vite umane, per il quale conia il termine di *motherism*sul modello dell'italiano "mammismo". Mentre infatti nelle società precedenti l'elevata numerosità dei figli e le precarie condizioni di vita rendevano la morte di uno o più di essi un evento frequente, in quella contemporanea il drastico calo delle nascite rende per una madre e un padre inconcepibile l'idea di poter perdere un figlio. Ciò perfino in forze armate formate da soli volontari, dove i genitori possono essere favorevoli all'arruolamento del figlio, ma poi ne vivono l'eventuale decesso "come uno scandalo oltraggioso, piuttosto che come un rischio professionale" (Luttwak, 1994, p. 25). È così che, secondo questo esperto di questioni strategiche, in un prossimo futuro i paesi che presentano una ridotta natalità non saranno più in grado di giocare un ruolo di grande potenza capace di proiettare la forza sulla scena mondiale laddove è necessario.

Diversa ma non incompatibile con quella appena esposta, è da ricordare

l'interpretazione micro-sociale del britannico Anthony King, ricavata dall'esperienza dei militari britannici caduti in Afghanistan. Tramontata l'etica patriottica e collettiva che aveva sostenuto il guerriero e sacralizzato il suo sacrificio – da Sparta fino alla prima guerra mondiale (culto del milite ignoto) – nell'iperindividualista società postmoderna l'elaborazione del lutto è ben più ardua che in passato e segue una procedura complessa. Lo studio dei necrologi e delle commemorazioni nel Regno Unito mostra come questa si articoli in due fasi. Nella prima il caduto viene recuperato nella sua identità di militare, vista tuttavia non più come l'adempimento di un dovere bensì come l'adempimento di una vocazione perseguìta con consapevolezza e coerenza. Nella seconda fase, il caduto viene presentato come persona, come figlio, coniuge, padre, "uno di noi" nella vita della comunità. Questo "addomesticamento" del soldato va di pari passo con una momentanea "svilirizzazione" dello Stato che, a differenza di quanto era accaduto per circa 2/3 del Ventesimo secolo, adesso "non è più libero di impiegare e perdere le proprie forze armate come crede [...] [avendo invece] il dovere di prendersi cura dei propri soldati, la cui protezione è ora (anche più che il compimento della missione) la sua responsabilità principale" (King 2010, p. 21).

Passando dall'analisi del nodo opinione pubblica/evento bellico che provoca vittime, alle "soluzioni" approntate dai governi per gestirlo, queste sono per lo più banalmente situazionali e, come tali, di corto respiro. La più grossolana è quella di impedire la circolazione dell'informazione sui costi umani del conflitto nel quale è stato coinvolto il paese, proibendo la pubblicazione delle notizie in materia e colpendo duramente i giornalisti colpevoli di occuparsene. Il silenzio è il metodo più utilizzato da una “democratura” come la Russia post-comunista. Qui i governi che si sono succeduti da Boris Eltsin in poi hanno steso la coltre del segreto su campagne repressive come quella condotta in Cecenia, la cui denuncia è costata la vita nel 2006 alla giornalista Anna Politovskaja. Efficace nel breve periodo, peraltro, la censura parziale o integrale non impedisce mai del tutto la circolazione dell'informazione, come persino l'allora Unione sovietica dovette constatare in occasione dell'occupazione dell'Afghanistan. Gli altissimi costi in vite umane (quasi un altro Vietnam) che comportò l'ultima delle guerre dell'URSS (1979-1989) ebbero un ruolo molto rilevante nella delegittimazione e nella crisi dell'intero regime sovietico.

Più morbida ma non meno insidiosa la strategia dei governi americani nell'affrontare i lutti delle campagne militari all'estero. L'esperienza del Vietnam era stata così traumatica per l'opinione pubblica, da ricevere una specifica denominazione giornalistica, la *body-bagsyndrome* (cioè la “sindrome dell'involucro di plastica” con il quale venivano reimbarcate per gli Stati Uniti le salme dei caduti): mai per tutta la durata della guerra in Iraq, né il presidente, né alcun membro del governo degli Stati Uniti si è mai recato a ricevere il corpo di uno di questi soldati. Relativamente facile da praticare in società tradizionali e rette da governi autoritari, la strategia del silenzio resta un semplice attenuatore dei fattori di crisi, quali appunto i costi umani delle

guerre, in un paese come gli USA, che della libertà di parola ha fatto (si pensi al ben noto II emendamento) un caposaldo costituzionale e, soprattutto, un valore condiviso e un comportamento diffuso. Su tale libertà i mass media "classici" hanno fondato il loro impero che, pur ridimensionato oggi dall'incalzare dei social, ha trovato in questi ultimi un erede non meno aggressivo.

Alla ricerca di soluzioni strutturali, non semplici palliativi bensì rimedi validi una volta per tutte, gli *spin doctor* della Difesa hanno messo a punto una soluzione creativa: *la guerra a vittime zero.* Nella narrazione che è andata sviluppandosi nell'establishment politico-militare negli ultimi decenni, la risposta all'eccessiva sensibilità del pubblico in tema di morti in guerra è promettere che questa sarà gratis o almeno che costerà molto poco in termini di caduti. Una promessa del tutto aleatoria, come mostrano le campagne combattute dagli Stati Uniti dopo il Vietnam. Queste infatti oscillano tra il record della guerra del Kossovo nel 1999 (appena due americani deceduti, causa incidente), la soddisfacente performance della guerra del Golfo 1 per la liberazione del Kuwait nel 1991 ("solo" 380 morti nei ranghi degli americani) sino ai ben più realistici (e drammatici) standard di conflitti a bassa intensità ma di lunga durata, quali l'Afghanistan (3.573 caduti dal 2001 a oggi) e l'Iraq (4891, dal 2003).

È evidente che l'impiego di armi avanzate e micidiali come quelle autonome segnerebbe un mutamento di paradigma nella conduzione dei conflitti, con irreparabili conseguenze sul piano etico, giuridico e comportamentale, oltre che operativo. In questo come in altri casi, l'unica misura efficace può essere la prevenzione attraverso la messa al bando delle LAWs, in quanto l'introduzione delle medesime, in una qualsiasi loro versione, innescherebbe un processo imitativo inarrestabile.

C'è da sperare che da parte della specie umana prevalga l'istinto di conservazione e non quello della massimizzazione della funzione di utilità, ancora più esiziale quando si sposta dal mercato all'uso della violenza politica. Da anni economisti e sociologi si interrogano sulla fine del lavoro umano (Rifkin, 2002), spiazzato da quello delle macchine, in particolare da quello delle macchine "intelligenti" (Ford, 2015). Come osservava Herbert Simon (1985, cit. in Veltri, 2018, p. 534), "che siano gli uomini o le macchine ad essere impiegati in un particolare processo, non dipende semplicemente dalla produttività relativa in termini fisici, ma altresì dal loro costo. E il costo dipende dal prezzo". Poiché quello della vita umana è il prezzo più elevato da pagare, il paradosso sarebbe che per risparmiare le proprie risorse umane si consentisse a una macchina di sacrificare quelle degli altri. Innegabilmente in guerra i dispositivi automatici "costeranno" di meno e quindi la tendenza "economica" sarebbe quella di farvi ricorso. Tuttavia, se in determinate circostanze un contingente di robot non si contrapponesse a un contingente di altri robot (come nella situazione simmetrica dal forte al forte), bensì ad esseri umani (nella situazione asimmetrica dal forte al debole),

lo scenario diventerebbe apocalittico[3].

[3] Si può notare per inciso che il differenziale tra il valore che la vita umana riveste in occidente e quello che riveste in oriente rappresenta un serio fattore di inferiorità di fronte a un particolare tipo di minaccia, quella terroristica, che il primo sta tentando di compensare con il parossismo tecnologico. Tipico, in questo caso, il pervasivo ricorso ai droni nelle "esecuzioni mirate" dei veri o presunti terroristi (IRIAD, 2017).

Bibliografia

Abrahamsson B., K. Ydén (2005). "Organizations, Co-ordinated Actions and Network Based Defence". In Ydén (eds.) *Directions in Military Organizing.* Stockholm: Försvarhögskolan: 169-184.

Ammendola T., Farina F. (1999). "Il gender gap dentro e fuori l'organizzazione militare". In Marletti C. (a cura di) *Politica e società in Italia*. Vol. II. Milano: Angeli: 755-775.

Bauman Z. (1995). *Il teatro dell'immortalità. Mortalità, immortalità e altre strategie di vita*. Tr. it. Bologna: il Mulino.

Burk J. (1998). "Riflessioni sulla fine della guerra fredda". In Id.*, La guerra e il militare nel nuovo sistema internazionale*. Tr. it. Milano: Angeli, pp. 27-49.

Douhet G. (1921). "Il dominio dell'aria e altri scritti"(a cura di L. Bozzo). In *Aeronautica militare*, Roma: Ufficio storico, 2002.

Erasmo da Rotterdam (2002). *Adagia 1530-1536*. Tr. it. (a cura di) D. Canfora, Roma: Salerno editrice.

Ford M. (2015). *Rise of the Robots: Technology and the Threat of a Jobless Future*. Tr. it. *Il futuro senza lavoro. Accelerazione tecnologica e macchine intelligenti*. Milano: Il Saggiatore, 2017.

Gorer G. (1965). *Death, grief and mourning in contemporary Britain*. London: Crescent Press.

Kagan R. (2004). *America's Crisis of Legitimacy, Foreign Affairs*, 82 (2): 65, 2004. Tr. it. *Il diritto di fare la guerra. Il potere americano e la crisi di legittimità*. Milano: Mondadori.

King A. (2010). "The Afghan War and 'postmodern' memory: commemoration and the dead of Helmand". In *The British Journal of Sociology*. Vol. 61, 1: 1-25.

Istituto di Ricerche Internazionali Archivio Disarmo - IRIAD (2017). "Droni militari: proliferazione o controllo". In *Sistema informativo a Schede*, n. febbraio-marzo.

Luttwak E.N. (1994). "Where are the great powers? At home with the kids". In *Foreign Affairs*, 73, n. 4, July-August: 23-28.

Nye J.S. (2005). "Soft Power: The Means To Success In World Politics". In *Public Affairs*, New York, 2004. Tr. it. *Soft power. Un nuovo futuro per l'America*. Torino: Einaudi.

Rifkin J. (1995). *The End of Work: The Decline of the Global Labor Force and the Dawn of the Post-Market Era*. Tr. it. *La fine del lavoro*. Milano: Mondadori, 2014.

Simon H. (1985). "The Corporation: will it be managed by machines?", in M. Anshen e

G.L. Bach (eds.), *Management and Corporations.* New York: McGraw Hill: 17-55.

Veltri F. (2018). "Dalla piramide alla clessidra. Verso una nuova divisione del lavoro sociale?". In A. Cipriani, A. Gramolati, G. Mari (a cura di), *Il lavoro 4.0. La Quarta rivoluzione industriale e le trasformazioni delle attività lavorative*. Firenze: Firenze University Press.

Westmoreland W.C. (1969). *Address by general W.C. Westmoreland, chief of staff,* U.S. Army, Washington, D.C., 14 October 1969, Congressional Record, U.S. Senate, 16 October 1969.

Cap. 1
Aspetti filosofici, etici e sociali delle armi autonome

1.1 Le fonti normative del dibattito etico sulle armi autonome

Il dibattito etico intorno alla progettazione, alla produzione, al dispiegamento e all'utilizzazione delle armi autonome si è principalmente sviluppato nella cornice offerta da due tra le principali impostazioni teoriche dell'etica normativa: l'etica dei doveri (detta anche etica deontologica) e l'etica delle conseguenze (detta anche etica consequenzialista) (Tamburrini, 2016; Tamburrini, 2020, cap. 4).

L'etica dei doveri si interroga su quali siano gli obblighi morali da adottare per giudicare il valore morale di azioni già compiute oppure per guidare decisioni e azioni, sia individuali sia collettive. Nel caso specifico delle armi autonome, l'etica dei doveri porta a chiedersi se vi siano degli obblighi morali per gli attori coinvolti in un conflitto bellico, il rispetto dei quali è minacciato dall'autonomia crescente o totale dei sistemi d'arma.

L'etica delle conseguenze si focalizza invece sui criteri per distinguere tra conseguenze moralmente buone e conseguenze moralmente cattive di un'azione, prescrivendo inoltre di giudicare il valore morale di un'azione solo in base a un bilancio delle sue conseguenze, reali oppure attese che siano. Nel caso specifico delle armi autonome, l'etica delle conseguenze porta ad interrogarsi sul bilancio complessivo tra le conseguenze moralmente buone e le conseguenze moralmente cattive che sono attese dallo sviluppo e dall'impiego delle armi autonome.

Di seguito sarà fornita una sintesi delle principali argomentazioni che vertono sull'ammissibilità morale delle armi autonome. Le principali obiezioni di natura morale che riguardano in modo specifico le armi autonome – rispetto ad altri sistemi d'arma che autonomi non sono – sono obiezioni di carattere deontologico emerse nel contesto teorico dell'etica dei doveri. Proprio in ragione della loro specificità partiremo, in questa esposizione sintetica, da un esame delle principali obiezioni di carattere deontologico, sollevate in base all'etica dei doveri, a proposito dell'autonomia operativa dei sistemi d'arma e dell'esercizio, da parte di una macchina, delle funzioni moralmente critiche di selezione degli obiettivi da attaccare e di esecuzione dell'attacco. Tali obiezioni deontologiche riguardano il rispetto di obblighi morali che incombono su tutti gli attori di un conflitto armato tra stati. Più specificamente, è stato sostenuto che le armi autonome pongono serie minacce per:

(i) Il rispetto delle norme del diritto internazionale umanitario;
(ii) il mantenimento della catena delle responsabilità nelle azioni belliche;
(iii) l rispetto della dignità degli esseri umani coinvolti nelle azioni belliche.

Procediamo a discutere nell'ordine i punti (i)-(iii), considerando innanzitutto le

minacce poste dalle armi autonome al rispetto del diritto internazionale umanitario.

1.2 Autonomia operativa e principi di distinzione e proporzionalità

Due sono le principali norme del diritto internazionale umanitario richiamate nella discussione etica e giuridica sulle armi autonome: il principio di distinzione e il principio di proporzionalità (codificati nel 1977 con i protocolli aggiuntivi alle Convenzioni di Ginevra del 1949; https://www.icrc.org/eng/assets/files/other/icrc_002_0321.pdf). Il principio di distinzione impone di limitare strettamente gli attacchi agli 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 da un lato e i nemici fuori combattimento o la popolazione civile inerme da un altro lato; di riconoscere e salvaguardare il personale militare sanitario e religioso, le unità sanitarie e i mezzi di trasporto civili o militari.

Consideriamo il problema di riconoscere un nemico fuori combattimento dal punto di vista delle tecnologie robotiche e dell'IA che giocano un ruolo cruciale nello sviluppo delle armi autonome. È da classificare come fuori combattimento un nemico che si astiene da ogni tentativo di fuga e soddisfa almeno una delle seguenti condizioni: a) si trova nelle mani di una delle parti avverse; b) esprime chiaramente l'intenzione di arrendersi; c) è in uno stato di assenza di coscienza oppure è stato neutralizzato da ferite o malattie che lo rendono incapace di difendersi (Protocollo I, art. 41).

Sviluppare dei sistemi che sappiano discriminare nel contesto delle operazioni belliche se una persona è fuori combattimento in base alle caratteristiche a)-c) non è un problema di facile soluzione per la ricerca in robotica e IA. Numerosi sono i problemi di ricerca che nella loro generalità rimangono ancora senza una soluzione adeguata. Come delimitare il compito percettivo di riconoscere quei comportamenti che esprimono chiaramente – ancorché in modi variegati e talora poco convenzionali – l'intenzione di arrendersi attraverso segnali, gesti del corpo o comunicazioni verbali? Come sviluppare un sistema percettivo artificiale capace di fare ciò senza incorrere in errori madornali, non potendo contare sulla condivisione tacita tra esseri umani di conoscenze e indizi ricavabili dal contesto culturale e antropologico?

Un confronto tra i migliori *benchmark* per i sistemi percettivi dell'IA e le prestazioni di un soldato competente, ben addestrato ed equipaggiato con sistemi tecnologici di supporto alla percezione (Tonin, 2019, p. 6) evidenzia come questi problemi non ammettono ancora una soluzione generale soddisfacente.

Le caratteristiche dei campi di battaglia sollevano ulteriori dubbi a proposito della capacità di un'arma autonoma di rispettare il principio di distinzione. È importante

sottolineare a questo riguardo che un campo di battaglia non è quell'ambiente ben ordinato, ripetitivo e privo di sorprese caratteristico della catena di montaggio robotizzata, dal quale sono state eliminate tutte le possibili fonti di perturbazione dell'azione robotica. Un campo di battaglia è un ambiente dinamico e poco strutturato, nel quale ogni attore sfida intenzionalmente la capacità di previsione degli avversari con iniziative e manovre a sorpresa, ivi inclusi tentativi di hackeraggio, *jamming* e *spoofing*. Questa caratteristica degli ambienti operativi delle armi autonome introduce delle difficoltà per la progettazione di test empirici che siano veramente adeguati a valutare la capacità delle armi autonome di affrontare correttamente le situazioni impreviste che possono insorgere sul campo di battaglia. Un comandante militare ben addestrato dovrà essere consapevole di questo problema e delle sue implicazioni sul grado di fiducia che si può riporre nella capacità di un'arma autonoma di rispettare le consegne nel corso delle azioni belliche.

I problemi appena ricordati di previsione del comportamento di un'arma autonoma nascono in considerazione delle caratteristiche e dell'evoluzione dinamica dell'ambiente esterno nel quale essa deve operare. Problemi previsionali altrettanto difficili nascono dalle caratteristiche delle componenti interne all'arma autonoma e dalle sue modalità di funzionamento. Caratteristiche interne che creano problemi previsionali rilevanti nascono in particolare dall'uso diffuso e difficilmente eliminabile delle tecnologie per l'apprendimento automatico (*machine learning*) nello sviluppo dei sistemi di percezione artificiale (vedi cap. 2).

Per rispettare il principio di distinzione, un'arma autonoma deve essere in grado di distinguere obiettivi militari legittimi da persone e cose che non lo sono. Per chiarezza espositiva, illustriamo il problema considerando due casi particolari e particolarmente calzanti. Il primo di questi riguarda il riconoscimento visivo di un pulmino scolastico e della sua conseguente esclusione dalla lista degli obiettivi militari in conformità con il principio di distinzione.

Da un test eseguito su un sistema percettivo dell'IA sono scaturiti risultati piuttosto allarmanti che riguardano direttamente questo problema di classificazione. Il sistema in questione, formato da una rete neurale profonda (in inglese *Deep neural network*), è stato addestrato con metodi di apprendimento automatico per questo tipo di reti a classificare delle immagini in ingresso come immagini di pulmino scolastico, di cane o di struzzo. Dopo la fase di addestramento, le reti neurali profonde presentano nel loro funzionamento a regime delle percentuali molto alte di classificazione corretta delle immagini in ingresso; ma possono anche incorrere in rari errori, che appaiono stravaganti e madornali rispetto alle capacità di classificazione percettiva di un essere umano.

Nei test eseguiti dopo la fase di apprendimento, l'immagine di un pulmino scolastico è stata inizialmente classificata in modo corretto dalla rete neurale profonda in questione. In seguito, sono state apportate piccole modifiche mirate alla stessa

immagine, applicando una metodologia di *adversarial testing* (Goodfellow*et al*., 2018) che ha lo scopo di mettere in luce le fragilità del sistema. Queste piccole modifiche sono impercettibili al sistema visivo umano integro, il quale continuerà a classificare correttamente la nuova immagine come quella di un pulmino scolastico. La rete neurale ha invece classificato come uno struzzo l'oggetto presente nella nuova immagine (Fig. 1, riga in alto).

Evidentemente, ai fini di una corretta applicazione del principio di distinzione, un pulmino scolastico ha una rilevanza maggiore di uno struzzo o di un altro animale. Risultati altrettanto sorprendenti, se valutati considerando la robustezza del sistema visivo umano alle medesime piccole perturbazioni, sono stati ottenuti su immagini di cani: modifiche impercettibili per l'occhio umano hanno nuovamente indotto il sistema a cambiare la classificazione iniziale corretta in quella di struzzo (Fig. 1, riga in basso).

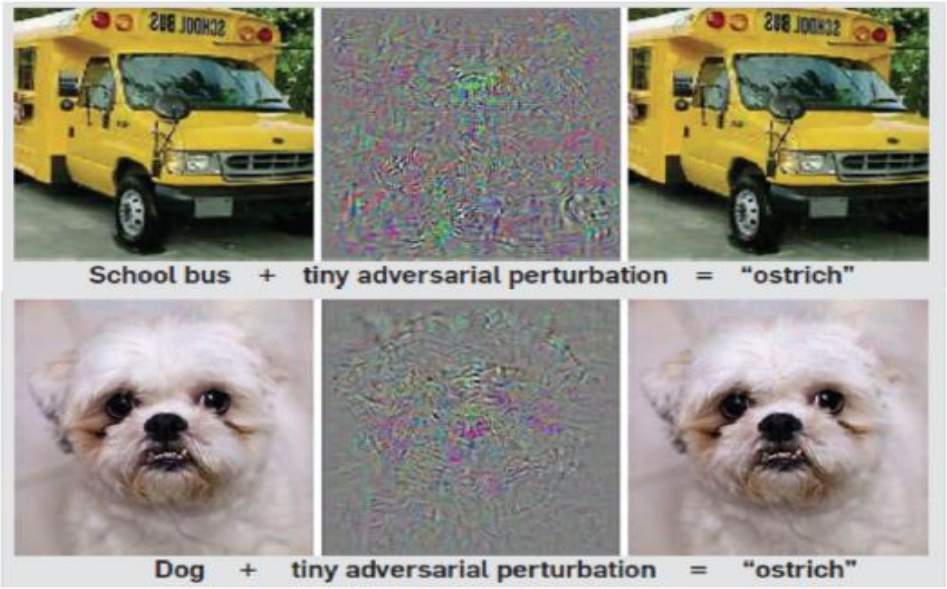


Fig.1. *Esempio di adversarial testing*. Fonte: Szegedi et al., 2013

Nella colonna di sinistra della Figura 1 le immagini sono classificate correttamente dal sistema; al centro la perturbazione introdotta (e ingrandita 10x per renderla visibile a occhio nudo) per passare dalle immagini della colonna di sinistra a quelle della colonna di destra; e infine, nella colonna di destra, le immagini così ottenute, che il sistema ha classificato come "struzzo".

Il secondo esempio di *adversarial testing* (Fig. 2), ovviamente rilevante per la discussione sulle armi autonome e il principio di distinzione, riguarda le piccole modifiche costruttive, del tutto ininfluenti per il sistema visivo umano, che hanno indotto un sistema percettivo dell'IA a classificare come un fucile automatico la riproduzione di una tartaruga ottenuta utilizzando una stampante 3D.

I modelli nei riquadri in rosso sono stati classificati come “fucile”; quelli nei riquadri in nero come “altro” rispetto alle classi “tartaruga” e “fucile”. Il sistema aveva ottenuto una percentuale vicina 100% di classificazioni corrette su modelli di tartaruga non modificati con tecniche di *adversarial testing*.

Fig. 2. *Adversarial testingsu modelli 3D di tartaruga.* Fonte: Athalye*et al*., 2018

Evidentemente, il sistema percettivo umano non è sensibile alle perturbazioni che hanno indotto le diverse classificazioni indicate nella Fig. 2. Ma queste possibilità di perturbazione del sistema artificiale che non possono ingannare il sistema visivo umano sollevano interrogativi importanti a proposito del rispetto del principio di distinzione da parte di un’arma autonoma. Ronald Arkin, un professore universitario di robotica che ha effettuato ricerche militari in ambito robotico, ha dichiarato senza mezzi termini che il dispiegamento e l’uso di armi autonome capaci di rispettare il principio di distinzione non è una meta raggiungibile nel breve periodo: “Vi sono profonde sfide tecnologiche ancora da risolvere, che riguardano tra l’altro la discriminazione efficace *in situ* degli obiettivi e il riconoscimento di chi è fuori combattimento.” (Arkin 2013 e 2015). In base a questa valutazione, Arkin ha proposto di introdurre una moratoria internazionale sulle armi autonome – e cioè una sospensione temporanea di qualsiasi attività finalizzata allo sviluppo, al dispiegamento e all’uso di armi autonome. Per Arkin, la moratoria dovrebbe rimanere in vigore fin quando non siano state raccolte prove sperimentali sufficienti per affermare, con un grado di fiducia sufficientemente elevato, che un’arma autonoma è in grado di rispettare il diritto internazionale umanitario almeno altrettanto bene di soldati competenti e ben addestrati.

Passiamo ora a considerare il principio di proporzionalità in relazione all’autonomia dei sistemi d’arma. Questo principio impone di non sferrare un attacco che abbia un costo atteso eccessivo – in termini di morti e feriti tra i civili o di danni a installazioni civili – rispetto al vantaggio militare concreto e diretto che ne potrebbe derivare (Protocollo I del 1977, art. 51.5.b). 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 può comportare l’uso di capacità cognitive ed emotive, di competenze sociali ed esperienziali che non sono alla

portata di un sistema dell'IA e della robotica. Esso richiede, tra l'altro, una valutazione degli effetti sul morale dei propri sottoposti o sul comportamento del nemico derivanti da un attacco che comporta il sacrificio di un gruppo di civili. Ma questo tipo di valutazione non è alla portata di un sistema dell'IA (Amoroso e Tamburrini, 2017). Allo stato dello sviluppo tecnologico attuale, si può prefigurare l'utilizzazione di un sistema dell'IA solo come strumento di supporto per valutazioni di proporzionalità che devono rimanere in capo ai comandanti militari.

In sintesi, di fronte alle richieste espresse dalle norme del diritto internazionale umanitario, il professore universitario di robotica Noel Sharkey ha osservato che "i robot autonomi o i sistemi dell'IA non hanno – né ora né in un futuro prevedibile – le proprietà che consentono di discriminare tra combattenti e civili o di prendere decisioni sulla proporzionalità degli attacchi" (Sharkey 2010, p. 378).

1.3 Armi autonome, responsabilità e dignità degli esseri umani

Le tecniche per lo sviluppo di test antagonistici menzionate nel paragrafo precedente rivelano errori percettivi commessi da un sistema dell'IA che potrebbero sfociare in un "disastro dell'IA": una decisione errata con gravi conseguenze dal punto di vista morale o legale. Un'arma autonoma che prende un pulmino scolastico per uno struzzo, o una tartaruga per un fucile automatico, potrebbe avere conseguenze materiali classificabili come crimini di guerra o crimini contro l'umanità se alla loro origine ci fossero degli esseri umani invece di una macchina.

Chi sarà ritenuto responsabile dell'attacco di un'arma autonoma al pulmino scolastico e della morte delle persone che si trovano a bordo? Escludiamo da subito l'arma autonoma: non essendo un'agente morale, non le si possono addossare responsabilità di alcun tipo per il verificarsi dell'evento in questione. Bisogna restringere la ricerca degli eventuali responsabili alle persone che hanno giocato un qualche ruolo nelle decisioni che hanno portato all'attivazione dell'arma autonoma e sono sfociate nell'attacco al pulmino scolastico. La lista delle persone coinvolte è piuttosto lunga. Vi sono gli ingegneri informatici, robotici e di altri settori che hanno contribuito a sviluppare l'arma autonoma; i responsabili della ditta produttrice; il personale politico che ha dato l'impulso all'acquisizione dell'arma autonoma da parte delle forze armate; i consulenti e i dipendenti del ministero della difesa che hanno stilato e approvato il piano di acquisizione; i militari incaricati dell'approvvigionamento dei sistemi d'arma, quelli che hanno stilato i manuali d'uso, il comandante in capo all'operazione militare, il personale militare incaricato di supervisionare il funzionamento dell'arma autonoma dopo la sua attivazione (Wagner, 2016). Più lunga è la lista, più facilmente la ricerca dei responsabili porterà al cosiddetto "problema delle molte mani". Ogni attore coinvolto in questi passaggi potrebbe concorrere in

qualche modo al verificarsi dell'evento, senza però che alcuni di essi possano essere considerati individualmente come responsabili di un crimine di guerra (Amoroso e Tamburrini, 2017; Bhuta e Pantazopoulos, 2016). Come è già accaduto in incidenti dovuti al malfunzionamento dei sistemi informatici (Nissenbaum, 1996, p. 25), si appura che numerose persone hanno concorso a vario titolo, ma si arriva a concludere che nessuna di esse ha dato un contributo davvero significativo al verificarsi dell'incidente. L'impiego delle armi autonome può perciò introdurre nuove difficoltà che riguardano la possibilità di attribuire responsabilità morali e penali nel caso di azioni compiute da un'arma autonoma che materialmente equivalgono a un crimine di guerra.

L'ultimo degli argomenti sviluppati contro le armi autonome nel quadro dell'etica dei doveri si basa sul rispetto della dignità umana (Amoroso et al., 2018). Questo argomento fornisce una motivazione razionale per la ripugnanza istintiva che suscita l'idea di affidare a una macchina il potere di vita o di morte su un essere umano (UNIDIR 2015, pp. 7-8). Il filosofo Peter Asaro ha sostenuto che dal rispetto della dignità umana discende il diritto di ogni essere umano di non essere deprivato arbitrariamente della vita. E perché una decisione di vita o di morte soddisfi questo requisito, ha affermato Asaro, essa deve essere presa da un altro essere umano, da qualcuno che possa provare empatia e avere compassione per chi è oggetto di una decisione di vita o di morte, piuttosto che da una macchina, la quale non ha la capacità di apprezzare il valore di una vita umana e di valutare adeguatamente il significato della sua perdita. Affidare a una macchina la decisione di togliere la vita a un essere umano, ha infine concluso Asaro, costituisce una violazione della dignità umana (Asaro, 2012).

L'argomento proposto da Asaro è stato efficacemente riformulato da Christof Heyns, in un rapporto stilato nella sua qualità di Relatore speciale delle Nazioni Unite per le esecuzioni extragiudiziarie, sommarie o arbitrarie. Heyns passa dalla prospettiva dell'arma che prende una decisione di vita o di morte alla prospettiva della vittima potenziale che la subisce. Quest'ultima non ha la possibilità di fare appello all'umanità condivisa di qualcuno che si trovi dall'altra parte. Il riconoscimento del valore intrinseco di un essere umano (della sua dignità in quanto uomo) sarebbe perciò negato a priori, se si affida il suo destino alla discrezionalità di una macchina (Heyns, 2016).

1.4 Armi autonome ed etica delle conseguenze

A questi vari argomenti contro l'autonomia dei sistemi d'arma che sono basati sull'etica dei doveri, si sono affiancate valutazioni morali delle armi autonome che sono invece basate sull'etica delle conseguenze. Dalla prospettiva teorica del consequenzialismo sono stati infatti sviluppati argomenti sia a favore sia contro lo sviluppo e il dispiegamento delle armi autonome. Cominciamo dagli argomenti di

stampo consequenzialista a favore delle armi autonome. Tali argomenti fanno leva su considerazioni di tre tipi. Tali considerazioni riguardano possibili scenari futuri che presuppongono sviluppi nel lungo periodo delle tecnologie robotiche e dell'IA che vanno ben al di là dello stato dell'arte scientifico e tecnologico:

a. Innanzitutto, si ipotizza che le armi autonome potranno garantire maggiore precisione nell'attacco agli obiettivi militari in alcuni scenari bellici di un futuro nel quale avranno superato i test volti a dimostrare che esse hanno raggiunto capacità di giudizio e di classificazione percettiva uguali o superiori a quelle di soldati competenti e ben addestrati.
b. Inoltre, le armi autonome che avranno superato tali test non saranno soggette a perturbazioni emotive, al contrario di un comune soldato. Per questo motivo, esse potranno essere vantaggiosamente sostituite ai comuni soldati, poiché consentiranno di ridurre le vittime negli eserciti che le dispiegano e di evitare quelle violazioni del diritto internazionale umanitario dovute allo stress e alle dure prove emotive alle quali sono sottoposti i soldati sul campo di battaglia.
c. Tali armi autonome potranno infine essere programmate in conformità con regole d'ingaggio più restrittive di quelle pensate per un comune soldato, poiché la vita di un soldato ha un valore morale incomparabilmente più alto di quello attribuito alla ferraglia di un'arma autonoma. In definitiva, un'arma autonoma potrà essere programmata per lanciare i suoi attacchi solo in circostanze che non lasciano dubbi sulle intenzioni ostili del nemico.

Il peso morale delle considerazioni a-c non può essere sottovalutato da una prospettiva consequenzialista, poiché prefigura situazioni di riduzione delle vittime sul campo di battaglia, sia nei propri ranghi sia in quelli dei nemici e degli astanti innocenti. Ma bisogna tuttavia ribadire che tali considerazioni riguardano solo un futuro indeterminato, che si realizzerà se e quando lo sviluppo tecnologico consentirà di sviluppare armi autonome in grado di esercitare capacità di giudizio e di classificazione percettiva uguali o superiori a quelle di soldati competenti e ben addestrati. È ancora più importante rilevare, da una prospettiva consequenzialista, che l'orizzonte delle conseguenze attese esaminate in a-c è ristretto ai singoli campi di battaglia. Perché limitarsi a considerare tra le conseguenze attese l'eventuale riduzione delle vittime su ogni singolo campo di battaglia? E perché concentrarsi selettivamente su ciò che una singola arma autonoma può fare in termini di precisione e di contenimento del volume di fuoco? L'etica delle conseguenze non impone di restringere ai singoli campi di battaglia la valutazione morale alle conseguenze attese, e nemmeno di ignorare le interazioni di ciascuna arma autonoma con altri agenti umani o artificiali. Uno sguardo più ampio sulle conseguenze attese potrà tenere conto delle perturbazioni sul funzionamento corretto delle armi autonome, che possono essere indotte da interazioni con altri agenti umani o artificiali, degli incentivi a iniziare nuovi conflitti che le armi autonome potrebbero offrire, di come le armi autonome potrebbero influire sul

mantenimento della stabilità e della pace tra le nazioni, di una nuova corsa alle armi autonome e dei processi di destabilizzazione che esse potrebbero innescare su scala regionale o globale.

Noel Sharkey ha allargato a un orizzonte più ampio la valutazione delle conseguenze attese dal dispiegamento delle armi autonome. Un'eventuale riduzione di perdite nel proprio esercito può indebolire l'ostilità dell'opinione pubblica al coinvolgimento del proprio paese in un conflitto e incentivare i decisori politici a iniziare una guerra. Inoltre, la riduzione attesa di vittime nel proprio esercito e sui singoli campi di battaglia potrebbe essere controbilanciata dal numero più alto di vittime, proprie o altrui, che i nuovi conflitti facilitati dalle armi autonome potrebbero causare nel lungo periodo (Sharkey, 2010). L'aspettativa stessa di una riduzione delle perdite sul singolo campo di battaglia potrebbe essere vanificata da episodi di "fuoco amico" provocati dall'hackeraggio di un'arma autonoma, da cyberattacchi che investono l'infrastruttura informatica (vedi cap. 4), malfunzionamenti e da interazioni impreviste con altri sistemi d'arma. Infine, le armi autonome portano con sé la minaccia di un ritmo accelerato dei conflitti, che potrebbe diventare incompatibile con i tempi di reazione degli operatori umani e sfuggire alle loro capacità di controllo (Altmann e Sauer, 2017).

Le interazioni impreviste tra sistemi dell'IA hanno generato già situazioni critiche in altri settori, per esempio per il mercato finanziario, a partire dal "flash crash" del 6 maggio 2010, allorché l'indice Dow Jones della borsa valori di New York registrò un ribasso del 9% nel giro di pochi minuti. Algoritmi dello stesso tipo sono stati additati come le cause più probabili del ribasso sostanzioso e repentino del valore della sterlina inglese avvenuto nell'ottobre 2016. Possibilità analoghe di interazioni competitive incontrollate si estendono anche ai sistemi d'arma autonomi basati sulla robotica e l'IA, come ha vividamente evidenziato Noel Sharkey in un recente articolo per la rivista *Scientific American*, sfruttando un parallelo con gli esiti di interazioni competitive osservati nel caso di algoritmi progettati per modificare il prezzo delle merci in funzione della domanda:

"A clear example of algorithmic confrontation run amok was provided by two booksellers, bordeebook and profnath, on the Amazon Web site in April 2011. Usually the out-of-print 1992 book The Making of a Fly sold for around $50 plus $3.99 shipping. But every time bordeebook increased its price, so did profnath; that, in turn, increased bordeebook's price, and so on. Within a week bordeebook was selling the book for $23,698,655.93 plus $3.99 shipping before anyone noticed. Two simple and highly predictable computer algorithms went out of control because their clashing strategies were unknown to competing sellers. what could happen if the complex combat algorithms of two swarms of autonomous weapons interacted at high speed. Apart from the uncertainties introduced by gaming with adversarial images, jamming, spoofing, decoys and cyberattacks, one must contend with the impossibility of predicting the outcome when computer algorithms battle it out. It should be clear that

these weapons represent a very dangerous alteration in the nature of warfare. Accidental conflicts could break out so fast that commanders have no time to understand or respond to what their weapons are doing—leaving devastation in their wake" (Sharkey, 2020, pp. 56-57).

Jürgen Altmann ha ulteriormente ampliato l'inventario delle conseguenze attese dal dispiegamento delle armi autonome, passando dalla prospettiva dei singoli campi di battaglia e dei singoli stati coinvolti in un conflitto alla prospettiva degli equilibri geopolitici globali: "Il peso dei vantaggi militari ipotizzati nel breve periodo dovrebbe essere contrapposto alle conseguenze probabili di lungo periodo per la sicurezza nazionale e, in modo particolare, per quella internazionale." (Altmann 2013, p. 137). Poiché i sistemi della robotica e dell'IA hanno costi di sviluppo e produzione relativamente bassi rispetto ad altre armi convenzionali, una corsa alle armi autonome potrebbe coinvolgere un numero consistente di stati e avere un forte impatto destabilizzante sugli attuali equilibri militari. Per lo stesso motivo, è ragionevole prevedere la loro acquisizione sul mercato illegale delle armi da parte di organizzazioni terroristiche.

Bisogna infine osservare che le armi autonome potrebbero avere un effetto destabilizzante anche sulle relazioni tra potenze nucleari. La nave Sea Hunter (vedi cap. 2), così come altri mezzi sottomarini autonomi in via di sviluppo, potrebbero essere utilizzati per identificare, tracciare ed eventualmente attaccare i sottomarini del nemico armati di missili balistici con testate nucleari. Poiché la possibilità di risposta a un attacco nucleare si basa prevalentemente sui sottomarini armati di missili balistici con testate nucleari, le armi autonome potrebbero contribuire a erodere le basi della deterrenza nucleare basata sulla distruzione reciproca assicurata (nota come deterrenza MAD, dall'acronimo inglese di *Mutually Assured Destruction*).

In conclusione, è opportuno sottolineare che l'etica delle conseguenze fornisce, al contrario dell'etica dei doveri, un panorama di argomenti e posizioni che non sono univocamente indirizzate contro le armi autonome. Il dispiegamento futuro delle armi autonome è stato infatti sostenuto da una prospettiva consequenzialista indicando la possibilità di ridurre le vittime sui campi di battaglia nei quali opereranno delle armi autonome.

È tuttavia importante notare che tale possibilità dipende dai progressi delle tecnologie robotiche e dell'IA per conseguire i quali, come ha sottolineato Arkin, sussistono "profonde sfide tecnologiche ancora da risolvere, che riguardano tra l'altro la discriminazione efficace *in situ* degli obiettivi e il riconoscimento di chi è fuori combattimento" (Arkin, 2013; 2015). Ed è ancora più importante notare che gli argomenti consequenzialisti a favore delle armi autonome assumono implicitamente che esse non avranno un impatto significativo al di là dei singoli campi di battaglia. Abbandonando questa ipotesi poco plausibile, un bilancio allargato dei costi e dei benefici attesi dovrà prendere in considerazione anche l'aspettativa di un ricorso più

facile alle armi e di una conseguente moltiplicazione dei campi di battaglia, di hackeraggi malevoli, di nuovi incentivi a una nuova corsa dalle armi, di rischi consistenti di destabilizzazione regionale e globale, e perfino di un indebolimento dei tradizionali meccanismi di deterrenza nucleare basati sui meccanismi della deterrenza MAD (Tamburrini, 2016).

In definitiva, un bilancio allargato delle conseguenze attese rafforza gli argomenti sul diritto internazionale umanitario, sul mantenimento della catena di responsabilità e sul rispetto della dignità umana che dalla prospettiva dell'etica dei doveri sono stati portati per contrastare lo sviluppo, la produzione e l'impiego delle armi autonome. Come vedremo nel cap. 7, questi argomenti di carattere etico hanno giocato un ruolo molto significativo nelle discussioni diplomatiche e politiche che si sono sviluppate sulla scia della proposta, avanzata da un nutrito gruppo di ONG, di messa al bando delle armi autonome.

Bibliografia

Altmann J. (2013). “Arms control for armed uninhabited vehicles: an ethical issue”. In*Ethics and Information Technology*, 15: 137-152.

Altmann J., Sauer F. (2017). “Autonomous Weapon Systems and Strategic Stability”. In*Survival. Global Politics and Strategy*, 59: 117-142.

Amoroso, D., Sauer F., Sharkey N., Suchman L., Tamburrini G. (2018). *Autonomy in Weapon Systems. The Military Application of Artificial Intelligence as a Litmus Test for Germany’s New Foreign and Security Policy*. Berlino: Heinrich Böll Foundation.

Amoroso D., Tamburrini G. (2017). “The Ethical and Legal Case Against Autonomy in Weapons Systems”. In *Global Jurist*, 17(3): 1-20.

Arkin R.C. (2013). “Lethal autonomous systems and the plight of the non-combatant”. In *AISB Quarterly*, 137: 1-9.

Arkin R. C. (2015). “The case for banning autonomous weapons: counterpoint”. In *Communications of the ACM*, 58(12): 48-49.

Asaro P. (2012). “On banning autonomous weapon systems: human rights, automation, and the de-humanization of lethal decision-making”. In *International Review of the Red Cross*, 94:687-709.

Athalye A., Engstrom L, Ilyas A., Kwok K. (2018). “Synthetizing robust adversarial examples”. In *Proceedings of the 35th International Conference on Machine Learning*.Stoccolma: PMLR 80:284-93.

Bhuta N., Beck S., Geiss R., Liu H.-Y., Kress C.(eds.) (2016). *Autonomous weapons systems: law, ethics, policy*. Cambridge: Cambridge University Press.

Bhuta N., Pantazopoulos S.E. (2016). “Autonomy and uncertainty: increasingly autonomous weapons systems and the international legal regulation of risk in Autonomous Weapons Systems”. In Bhuta N. et al. (2016): 284-300.

Goodfellow I., McDaniel P., Papernot N. (2018). “Making machine learning robust against adversarial testing”. In *Communications of the ACM*, 61(7): 56-66.

Heyns C. (2016). “Autonomous weapons systems: living a dignified life and dying a dignified death”. In Bhuta et al. (eds.) (2016): 3-19.

Nissenbaum H. (1996). “Accountability in a computerized society”. In *Science and Engineering Ethics*, 2:25–42.

Sharkey N. (2010). “Saying ‘no!’ to lethal autonomous targeting”. In *Journal of Military Ethics*, 9: 369-383.

Sharkey N. (2020). “Autonomous Warfare”, *Scientific American*, February 2020: 52-57.

Szegedi C., Zaremba W., Sutskever I., Bruna J., Erhan D., Goodfellow I., Fergus R. (2013). “Intriguing properties of Neural Networks”, arXiv:1312.6199, Cornell University, Ithaca, (https://arxiv.org/abs/1312.6199).

Tamburrini G. (2016), “On banning autonomous weapons systems: from deontological to wide consequentialist reasons”, in Bhuta N. et al. (eds) (2016: 122-142.

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

Tonin M.(2019). *Artificial Intelligence: implications for NATO’s armed forces*. Rapporto pubblicato il 5 aprile 2019, NATO Parliamentary Assembly, Science and Technology Committee, Bruxelles, (https://www.nato-pa.int/document/ 2019-stctts-report-artificial-intelligence-tonin-088-stctts-19-e).

UNIDIR (2015). *The Weaponization of Increasingly Autonomous Technologies: Considering Ethics and Social Values*. UNIDIR Resources n. 7, United Nations Institute for Disarmament Research, New York, (https://www.unidir.org/ publication/weaponization-increasingly-autonomous-technologies-autonomous-weapon-systems-and-cyber).

Wagner M. (2016). “Autonomous Weapon Systems”. In Wolfrum R. (ed.), *Max Planck Encyclopedia of Public International Law*.Oxford: Oxford University Press.

Cap. 2
Caratteristiche, prospettive e problematicità dell'Intelligenza Artificiale

In questo capitolo viene fornita una breve introduzione all'Intelligenza Artificiale (IA), con particolare riferimento alle tecniche di apprendimento automatico (*Machine Learning*, ML, in inglese) e una discussione dei limiti e delle problematicità di queste ultime. Il paragrafo 2.1 fornisce una breve introduzione storica delle tecnologie dell'IA e cenni sulle sue principali applicazioni e sugli aspetti di natura più concettuali e scientifici della ricerca in IA. Il paragrafo 2.2 è focalizzato sulle tecniche di ML: vengono brevemente descritte le varie fasi di una tipica procedura di ML e alcune fra le tipologie di ML più diffuse, anche con l'aiuto di qualche esempio chiarificatore. Il paragrafo 2.3 mette in evidenza alcuni limiti e le conseguenti problematicità del ML, anche con riferimento al problema di attacchi informatici *specifici* per i sistemi di ML, ai quali ovviamente si aggiungono quelli degli attacchi cibernetici tradizionali perpetrabili sui computer e i programmi che implementano questi sistemi, oggetto del Capitolo 4 del presente documento. Infine, nel paragrafo 2.4. vengono presentate alcune conclusioni che si possono trarre, a partire da quanto descritto nei precedenti paragrafi, su questa tecnologia e sul suo uso odierno.

2.1 Introduzione

L'IA nasce ufficialmente nell'estate 1956 quando John McCarthy, uno dei padri di quella che lui stesso avrebbe chiamato *Artificial Intelligence*, all'epoca giovane assistente alla cattedra di matematica al Dartmouth College ad *Hanover* (Stati Uniti), organizzò il *Dartmouth Summer Research Project on Artificial Intelligence*, un *workshop* al quale parteciparono molti pionieri dell'IA, come Marvin Minsky, Herbert Simon e Allen Newell, e altri eminenti scienziati come Claude Shannon e Nethaniel Rochiester[4]. L'obiettivo del *workshop* era quello di chiarire e sviluppare concetti e idee sulle cosiddette *thinking machines*.

A tutt'oggi, però, non si è pervenuti ad una definizione soddisfacente di IA; in questo documento utilizziamo quella – in un certo senso riduttiva – proposta in UNIDIR (UNIDIR, 2018, p. 2), secondo cui "[l']intelligenza artificiale è il campo di studi dedicato a rendere intelligenti le macchine. L'intelligenza misura la capacità di un sistema di determinare la migliore linea d'azione per raggiungere i propri obiettivi in una vasta gamma di ambienti" (traduzione in italiano a cura degli autori).

[4] Si veda, ad esempio, (Franklin, 2014).

In pratica, molti ricercatori nel campo dell'IA, piuttosto che affrontare il problema di una definizione teorica completa della disciplina, preferiscono darne una caratterizzazione ottenuta affrontando, di volta in volta, specifici problemi per la soluzione dei quali il senso comune ci suggerisce che sia necessaria una certa dose di "intelligenza"–come ad esempio giocare a scacchi–e realizzare delle macchine capaci di risolverli autonomamente. In questo modo, l'*avanzamento delle conoscenze* di IA viene definito indirettamente, attraverso gli *sviluppi tecnologici* dell'IA. Così, nel 1952, un programma di IA è riuscito a vincere a *tris*, nel 1994 fu il turno della dama, quindi degli scacchi nel 1997, fino allo sviluppo, nel 2014, di programmi capaci di giocare e vincere sugli umani a vari giochi Atari, a Go (2016) e a poker (2017).

Un effetto secondario di questa tendenza è che solo le macchine più all'avanguardia vengono considerate "intelligenti": ad esempio, negli anni Sessanta venivano considerate intelligenti le macchine capaci di giocare a scacchi, mentre oggi queste macchine vengono considerate dei semplici programmi informatici più o meno efficienti. Con il passare degli anni, l'"intelligenza" delle macchine si è spostata venendo associata a compiti che noi riteniamo più complessi o creativi, come, per es. il gioco del Go. In altri termini, ciò che accade è che la capacità di svolgere un certo compito – per es. di risolvere un certo problema – viene considerata espressione di intelligenza solo fino a quando non viene realizzata una macchina capace di eseguirlo.

In ogni caso, le tecnologie dell'IA, e in particolare di quella branca dell'IA che va sotto il nome di *Machine Learning*, sono state applicate con notevole successo in vari domini, che vanno dal *riconoscimento* di immagini, della voce e del linguaggio naturale, alla *traduzione automatica* fra lingue (umane), all'*aggregazione* e *analisi* di immense quantità di dati (*data analytics* e *data science*) con fini *predittivi* o di *diagnosi* e al controllo di *sistemi autonomi*, come i veicoli a guida autonoma. Infine, sistemi di IA hanno sempre suscitato interesse in ambienti militari (Din, 1986); le applicazioni dell'IA in questo contesto possono andare dai sistemi C3IR (*Command, Control, Communication, Intelligence and Reconnaissance*) a quelli di *supporto alle decisioni*, fino alle *armi autonome*[5].

Concludiamo questa breve introduzione sottolineando che, accanto agli aspetti puramente tecnologici citati fino adesso e che costituiscono argomento di primario interesse nel contesto del presente documento, l'IA come disciplina squisitamente *scientifica* ha visto e vede tuttora sviluppi concettuali e teorici di altissimo livello, che vanno dalla creazione e studio di *modelli teorici* per la *rappresentazione della conoscenza*, per la *pianificazione* (*planning*) e l'*apprendimento* (sul quale ci soffermeremo più in dettaglio più avanti) e per la *explainability* dei procedimenti computazionali e logici che portano ai risultati generati da sistemi di IA, ai sistemi

[5] Si veda, fra gli altri: (DSB, 2016), (Allen et al., 2017), (Boulanin et al., 2017), (Cummings, 2017), (Dyndal et al. 2017), (Amoroso et al. 2018), (CRS, 2018), (UNIDIR, 2018), (Rossi, 2019).

logico-deduttivi per il ragionamento automatico, che fondano le proprie radici, anche storiche, nella logica matematica e nella teoria della calcolabilità; lo studio di questi ultimi ha permesso di sviluppare, ad esempio, dimostratori automatici di teoremi. Molte classi di logiche sviluppate o utilizzate nell'ambito della ricerca di base sull'IA come le logiche *epistemiche* (van Ditmarsch et al., 2015), *temporali* (Emerson, 1990), o *spaziali* (Aiello et al., 2007) sono specializzazioni della logica *modale* (van Benthem et al. 2006), un campo di proficua ricerca nell'ambito della logica matematica.

Infine, va citato lo studio e lo sviluppo della cosiddetta *Swarm Intelligence*, un paradigma utilizzato anche in ambito IA nel quale vengono sviluppati algoritmi ispirati al comportamento di popolazioni di agenti biologici, come colonie di insetti, stormi di uccelli o banchi di pesci. In questo caso, ogni singolo agente segue regole comportamentali e di interazione estremamente semplici che, però, danno origine ad interessanti *proprietà emergenti* della popolazione al punto tale che quest'ultima, nell'insieme, può risolvere in maniera estremamente efficace ed efficiente problemi di notevole complessità[6].

2.2 Machine Learning

Fra le tecniche sviluppate nell'ambito dell'IA per scopi civili, quelle di ML hanno riscosso particolare successo negli ultimi anni; nel resto di questo capitolo ci concentreremo quindi sul ML. Questo successo è in parte dovuto alla incredibile quantità di dati su svariati aspetti del problema di volta in volta di interesse e alla grande capacità di calcolo oggi disponibili. La disponibilità di dati è a sua volta conseguenza della diffusione di sensori disponibili a basso costo che possono monitorare praticamente qualunque aspetto fisico e sociale del pianeta e che vanno dai sensori *ad hoc* dedicati al monitoraggio di particolari fenomeni fisici, ai dispositivi elettronici di uso comune come gli *smart-phone*, i *tablet* e, con l'avvento dell'*Internet of Things*[7], tutti i più comuni elettrodomestici, i mezzi di trasporto e, più in generale, le infrastrutture (regionali, nazionali e urbane).

Il ML rappresenta un paradigma duale rispetto all'informatica tradizionale, ma anche all'IA "tradizionale" (a volte indicata dall'acronimo "GOFAI": *Good Old-Fashioned Artificial Intelligence*). Queste ultime, per definizione, hanno come obiettivo la soluzione di problemi che, per loro stessa natura, risultano difficili per gli umani, ma di facile automazione e quindi relativamente semplici per le macchine, ad esempio perché ripetitivi e basati su regole di inferenza logica. Viceversa, il ML è una disciplina che cerca di affrontare e risolvere quei problemi che, invece, non sono facilmente

[6] Si veda, ad esempio la rivista scientifica specializzata *Swarm Intelligence*, Dorigo, M. (Ed. in Chief). Berlin, D: Springer.

[7] Si veda, ad esempio (GAO, 2017).

descrivibili in termini di regole logiche, né necessariamente ripetitivi, e che, spesso, risultano "semplici" per gli umani, come, ad esempio, riconoscere gli oggetti che compongono una scena della vita comune, come osservato, fra gli altri, da Goodfellow (2016).

In altri termini, mentre le attività cognitive di alto livello, come ad esempio il ragionamento, possono essere, entro certi limiti, ricreate artificialmente utilizzando le tecniche tipiche della GOFAI, quelle di livello più basso, più assimilabili, per esempio, all'apprendimento umano dei primi anni di vita, sono più facilmente simulabili utilizzando tecniche di ML, a patto che si abbiano dati di alta qualità e potenza di calcolo sufficienti.

Il ML è una disciplina che vede integrati aspetti e risultati scientifici dell'informatica, della statistica e dell'algebra lineare, con intuizioni provenienti da altre discipline, come le neuroscienze. La caratteristica fondamentale delle tecniche ML è che, a differenza delle altre tecniche tipiche dell'informatica, esse affrontano il problema di programmare i computer a insegnare a loro stessi (e quindi a imparare) a partire dai dati e dall'esperienza (Buchanan et al., 2017).

Gli utilizzi principali del ML includono quelli di *classificazione*, quelli di *strutturazione* di dati originariamente non strutturati e di *riconoscimento* di *pattern*. Fra i principali domini di applicazione vanno citati: *computer vision*, cioè la capacità di un computer di riconoscere e identificare specifici oggetti in una immagine; elaborazione del linguaggio naturale, come la traduzione automatica da una lingua ad un'altra; *Internet of Things*, dove i dispositivi di uso comune possono apprendere le abitudini e le limitazioni degli utilizzatori facilitandone quindi l'uso; supporto al *design* e alla *ricerca scientifica*, dove sistemi artificiali di supporto intelligenti possono (aiutare a) ideare soluzioni innovative; sistemi medicali, dove sottosistemi di IA–per esempio sistemi di riconoscimento di immagini–possono fornire un valido supporto sia in fase di diagnosi che in fase di definizione delle terapie –per esempio nella segmentazione propedeutica alla radioterapia; sistemi di trasporto, sia per quanto riguarda gli aspetti di controllo del traffico che in relazione allo sviluppo di veicoli a guida autonoma; sistemi di supporto ai procedimenti giudiziari, come quelli di *polizia predittiva*; sistemi militari, sia come (sottosistemi di) sistemi di supporto alle decisioni che come componenti di armi autonome.

Le *fasi* di una tipica procedura di ML sono quattro:

1) programmazione, cioè la definizione, progettazione e implementazione di una procedura *software*[8] con la quale si istruisce un computer a "imparare" a partire da un insieme di dati, il cosiddetto *training data set*;

[8] Naturalmente, spesso è possibile sostituire questa fase con una di semplice "acquisizione" del software, sia dal mercato che attraverso l'uso di prodotti *freeware*.

2) allenamento (*training*), che consiste nell'esecuzione del programma di cui al punto precedente su un *training data set* al fine di impostare *e calibrare* i (molti!) parametri presenti nel programma; il *training data set* è costituito da dati che possono – ma non necessariamente devono – essere "etichettati", come vedremo più avanti;
3) *testing* del programma calibrato al punto precedente, e cioè esecuzione *controllata* dello stesso, su *un altro* insieme di dati – il *testing data set*– e relativa analisi dei risultati al fine di valutarne la qualità;
4) e infine, se la fase (3) viene completata con una valutazione positiva, uso del programma calibrato su *nuovi* dati (diversi da quelli di *training* e da quelli di *testing*).

Se la fase (3) non si conclude positivamente, si dovrà procedere a ulteriori fasi di *training* ed eventualmente rivedere anche le scelte di progettazione o acquisizione adottate nella fase di programmazione. Va notato subito che, a seconda del tipo di approccio al ML che si è scelto, è possibile che attività di *training* abbiano luogo anche durante l'uso del programma, come vedremo nel seguito.

Si possono identificare vari tipi di ML, fra i quali il *supervised* ML, l'*unsupervised* ML, il *reinforcement learning* e il *deep learning.* Caratteristica principale del *supervised* ML è il fatto che il *training data set* è costituito da dati *etichettati*: molto semplicemente, questo significa che ogni dato nel *data set* è accompagnato da un'etichetta che lo descrive e che, di norma, è creata *manualmente*. Facciamo un semplice esempio: uno degli usi del ML è quello del riconoscimento di testi scritti a mano (e la loro trascrizione in testo basato su un insieme di caratteri standard). Se, per semplicità, ci limitiamo al riconoscimento di cifre numeriche da 0 a 9 scritte a mano, possiamo immaginare il *training data set* come una *grande* tabella, ogni riga della quale è costituita da due campi: uno contiene un'immagine di una cifra manoscritta e l'altro l'etichetta, cioè la sua trascrizione; in Figura 2.1 riportiamo una infima parte di una tale tabella, con due immagini di altrettanti modi diversi di scrivere il numero 3, ma entrambi ovviamente associati al carattere "3":

Immagine	Etichetta
	3
	3

Fig. 2.1. *Un frammento di training data set per il riconoscimento delle cifre scritte a mano*

Il *training data set* viene usato dal programma di ML per impostare e, soprattutto, calibrare i *parametri* caratteristici del particolare *modello computazionale* di ML selezionato per l'utilizzo nell'applicazione d'interesse. Esistono svariati *modelli computazionali* per il ML *supervised*, che vanno dalla semplice regressione lineare, alle

Support Vector Machines alle reti neurali, solo per citarne alcuni (Goodfellow et al. 2016), ma caratteristica comune a *tutti* questi modelli computazionali è il fatto di essere fortemente *parametrizzati*; il cosiddetto "*training*" consiste nell'esecuzione di procedure, tipicamente iterative, per fissare i valori di questi parametri e calibrare questi ultimi per affinamenti successivi. La calibratura dei parametri deve evidentemente essere guidata da un qualche criterio che il programma di ML deve seguire; questo criterio è rappresentato da una *funzione di costo* da minimizzare; quest'ultima, in buona sostanza, rappresenta l'errore che, a seguito di una certa impostazione dei parametri, la macchina può commettere nel classificare o riconoscere i dati e che, quindi, va minimizzato. In ultima analisi, quindi, l'operazione di *training* altro non è che la soluzione di un problema di *minimizzazione* di una certa funzione matematica.

Implicito in quanto appena detto è il fatto che si abbiano rappresentazioni *numeriche* degli oggetti sui quali si fa ML. Ad esempio, nel caso di riconoscimento di oggetti in immagini, per semplicità in bianco e nero, la rappresentazione digitale dell'immagine è una matrice numerica bidimensionale di *pixel*, ciascuno dei quali è portatore di un valore numerico che identifica l'intensità luminosa del *pixel* stesso; questa matrice, a sua volta, può essere rappresentata come un vettore numerico – cioè una sequenza di numeri – di dimensione pari al numero totale di *pixel* dell'immagine. Solo a titolo di esempio, in Figura 2.2 riportiamo una piccola matrice di (4 x 4 =) 16 *pixel* (nei sistemi reali, si hanno matrici con migliaia o milioni di *pixel*!) e la sua rappresentazione come vettore, dove le righe della matrice sono semplicemente giustapposte, una a fianco dell'altra. In realtà, il vettore contiene dei numeri, che vanno da 0, corrispondente al nero (intensità minima), a 255, corrispondente al bianco (intensità massima), come rappresentato dal vettore numerico collocato nella parte superiore della figura.

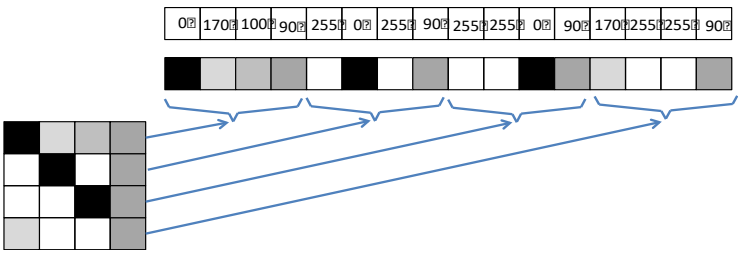
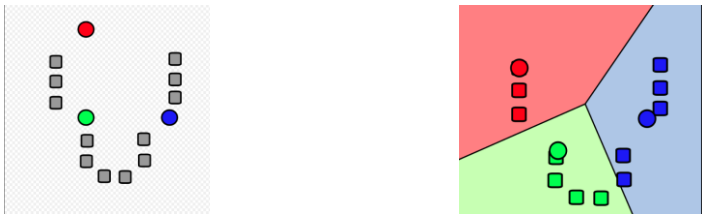


Fig. 2.2. *Esempio di rappresentazione numerica di immagini*

Quindi, un'immagine digitale è rappresentata da un vettore numerico. Esso rappresenta l'input del modello computazionale, il cui output può a sua volta essere un vettore, ad esempio con tanti elementi quanti sono gli oggetti diversi che si vogliono identificare: ogni elemento fornisce la probabilità che una data immagine in input contenga un certo specifico oggetto.

Il ML *unsupervised* viene usato tipicamente quando non si ha a disposizione dei *training data set* etichettati. In questo caso, sono gli algoritmi stessi che, ricevendo come input un (grande) insieme di dati non strutturato, cercano di identificare, in esso, dei *pattern*, una qualche struttura. Un esempio è il metodo *k-means clustering*: dato uno spazio multidimensionale (che rappresenta oggetti di interesse), l'algoritmo lo partiziona in *k* sottoinsiemi, ciascuno rappresentato dal suo centroide[9] e in modo tale che sia minimizzata la distanza fra ogni punto e il centroide del suo sottoinsieme di appartenenza (Buchanan et al. 2017). La Figura 2.3 fa vedere un esempio di applicazione dell'algoritmo di k-means, per k=3: in questo caso, vengono inizialmente scelti (di solito in modo causale) 3 punti, rappresentati dai tre tondini colorati in rosso, verde e blue nella Figura 2.3(a). L'algoritmo, procedendo in modo iterativo produce una partizione dell'insieme di tutti i punti (tondini e quadratini) secondo il criterio sopra specificato, mostrata in Figura 2.3(b); come si può facilmente notare, ogni punto di un dato colore è più vicino al centroide (tondino) dello stesso colore di quanto non lo sia ai centroidi di colore diverso.

Questa tecnica consente quindi di aggregare i dati di un certo insieme attorno ai *k*centroidi e quindi può risultare utile in quelle applicazioni che richiedono classificazioni dei dati come, ad esempio, l'identificazione di cellule cancerogene in un campione di tessuto, il *clustering* di parole con significati simili, le analisi di mercato o addirittura l'identificazione di mine in un campo di battaglia (Landman et al. 2019).



(a) (b)

Fig. 2.3. *Esempio di applicazione dell'algoritmo di k-means, con k=3.* Fonte: www.wikipedia.org[10]

Infine, uno degli usi più comuni del ML *unsupervised* è quello di scoprire una qualche struttura in insiemi di dati per poter poi utilizzare la conoscenza acquisita sulla struttura nella progettazione di sistemi di ML *supervised.*

Nel caso del *reinforcement learning* un agente artificiale (un programma in

[9] Il *centroide* di un insieme di punti in uno spazio multidimensionale è il punto che rappresenta la *media* dei punti nell'insieme; in altri termini, ogni componente del centroide è la media dei valori della stessa componente dei punti dell'insieme.

[10] https://en.wikipedia.org/wiki/K-means_clustering. By I, Weston.pace [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0/)] - accesso effettuato il 16 ottobre 2019.

esecuzione su uno o più computer) può "imparare" semplicemente interagendo con il suo ambiente. Tipicamente l'agente esegue un'azione, in maniera conforme a certe regole, ne "osserva" (tramite opportuni sensori) l'effetto sull'ambiente e quindi determina se l'azione è stata di aiuto per il raggiungimento degli obiettivi dell'agente. L'ambiente può anche essere costituito dall'agente stesso: è questo il metodo utilizzato da *Alpha Zero*per imparare a giocare a *Go* e ad altri giochi (Buchanan, 2017), (UNIDIR, 2018).

Infine, il *deep Learning* può essere pensato, in termini generali, come un insieme di tecniche "architetturali" attraverso le quali vengono combinate varie tipologie di ML, sia *supervised* che *unsupervised*, per l'estrazione di caratteristiche rilevanti dei dati in input e per l'apprendimento attraverso livelli multipli di rappresentazione che corrispondono a diversi livelli di astrazione che quindi formano una gerarchia di concetti (Wikipedia, 2018).

Ad esempio, il *deep learning* può combinare un processo *unsupervised* per apprendere le caratteristiche dei dati sottostanti, come i bordi di una faccia, e quindi fornire tali informazioni a un algoritmo di apprendimento *supervised* per riconoscere le caratteristiche e produrre il risultato finale, come identificare correttamente una persona in una foto (Buchanan, 2017).

2.3 Limiti e problematicità del Machine Learning

È innanzitutto opportuno premettere che i sistemi di IA, inclusi quindi quelli di ML, in ultima analisi sono costituiti da programmi in esecuzione in computer progettati e costruiti con tecnologie sostanzialmente tradizionali[11]. Di conseguenza, i sistemi di IA sono vulnerabili a tutti gli attacchi cibernetici che sfruttano vulnerabilità dei normali sistemi informatici, come verrà meglio descritto nel Capitolo 4 di questo documento.

Detto questo, bisogna sottolineare il fatto che il comportamento di un sistema di ML (*supervised*) durante la fase di uso dipende fortemente dalla *qualità* dei dati *e* delle procedure di *training*. Fra le altre cose, va considerato il fatto che da tale qualità dipende anche la possibilità di evitare che la macchina, durante l'uso, esibisca dei pregiudizi (*bias*, in inglese).

La creazione di *bias* nella fase di *training* può dipendere dal fatto che il sistema di ML abbia eseguito un addestramento poco accurato, cioè su dati che non rappresentano bene la popolazione di oggetti sui quali la macchina viene usata; oppure dal fatto che il *training* sia stato sì accurato, magari anche molto accurato, ma su dati a loro volta *biased*; oppure, semplicemente, dal fatto che il sistema non sia stato

[11] Va però detto che cominciano ad essere prodotti dei dispositivi hardware progettati e sviluppati appositamente per applicazioni di IA come il *Tensor Processing Unit*, un circuito integrato per reti neurali di Google (Wikipedia, 2019).

sufficientemente addestrato (Buchanan, 2017).

La presenza di *bias* è un problema particolarmente insidioso. “Nel peggiore dei casi, l'apprendimento automatico può nascondere discriminazioni con l'imprimatur della scienza”. “Ad esempio [...], una scuola di medicina britannica ha usato un algoritmo che scartava candidati qualificati di sesso femminile o provenienti da minoranze perché era stato addestrato sulle decisioni prese in precedenza da una commissione di valutazione non imparziale. Un'indagine di *ProPublica* ha rilevato che un algoritmo sviluppato dalla società *Northpointe, Inc*. per fornire un punteggio di valutazione del rischio [di supporto] per i giudici [nell'emissione di una] condanna era *biased* rispetto alla razza e inaccurato. Gli afroamericani avevano molte più probabilità di essere etichettati come "ad alto rischio", ma [...] gli afroamericani che erano stati etichettati come "ad alto rischio" avevano [di fatto] molte meno probabilità di commettere un altro crimine rispetto ai bianchi "ad alto rischio" (Buchanan, 2017, p. 32, traduzione in italiano a cura degli autori).

È quindi di estrema importanza poter disporre di *grosse* quantità di *dati* di *alta qualità* e di procedure di *training* altrettanto affidabili. Questo, non sempre è possibile e, soprattutto, è particolarmente difficile per domini di applicazione collegati al campo di battaglia, soggetto, per definizione alla *fog of war*.

Inoltre, riteniamo di fondamentale importanza notare che, dato un certo problema, per esempio, di classificazione, diversi algoritmi e modelli computazionali possono dare risultati abbastanza diversi. Si consideri, ad esempio, la Figura 2.4 (a); in questo caso gli oggetti in input sono rappresentati da punti di uno spazio bidimensionale, cioè coppie di numeri. Questo significa semplicemente che stiamo considerando degli oggetti del mondo reale dei quali riteniamo importanti soltanto due caratteristiche. Sulla base di queste caratteristiche, vogliamo classificare gli oggetti come "buoni" o meno. Ad esempio, i punti potrebbero rappresentare delle email (attraverso un paio di indicatori quali, per esempio, la lunghezza e il numero di parole di un certo vocabolario): nella figura quelli blue rappresentano email "buone", quelli rossi rappresentano invece degli *spam*. Possiamo eseguire il *training* di una rete neurale su un *data set* costituito da email che sappiamo essere "buone" e da altre che sappiamo essere *spam* e poi utilizzare la rete per classificare le nuove email, man mano che arrivano, come buone o *spam*.

La classificazione operata dalla rete neurale può essere rappresentata graficamente suddividendo il piano in varie zone, ciascuna corrispondente ad una data probabilità di avere a che fare con uno *spam,* come si vede in Figura 2.4(b): se un'email appena arrivata "casca" nella zona in alto a sinistra, allora la probabilità che sia uno spam è molto alta, mentre se va a cadere in basso a destra questa probabilità è molto bassa; le zone intermedie corrispondono a valori intermedi di probabilità di *spam.* Ovviamente, tutto questo sempre *secondo la rete neurale*.

(a) (b)

Fig. 2.4. *Un insieme di dati di input e la loro classificazione da parte di una rete neurale.* Fonte: Pedregosa et al., 2011. Figure tratte da sci-kit-learn "Classifier Comparison"[12]

Infatti, cambiando modello/algoritmo, sebbene le aree blue e rossa grossolanamente si sovrappongono (nel senso che tendenzialmente il rosso sta in alto a sinistra e il blue sta in basso a destra), i dettagli della classificazione cambiano sensibilmente, come mostrato nella Figura 2.5, relativa sempre allo stesso insieme di input della Figura 2.4(a) e allo stesso *training data set*.

Questo semplice esempio fa comprendere come, mentre l'uso di tecniche di ML può costituire un utile *supporto* per il *decision-making*, dando delle indicazioni "di massima" su come, ad esempio, classificare oggetti, scenari o situazioni di interesse, affidarsi a tali tecniche come *unico strumento* decisionale possa risultare estremamente rischioso, per lo meno allo stato attuale dell'arte.

Inoltre, l'esempio fa capire, con assoluta chiarezza, che ogni particolare modello computazionale ha i suoi punti di forza e i suoi punti deboli; di conseguenza, i professionisti del ML devono spesso provare più di un modello/algoritmo per determinare quale, fra essi, risponde meglio alle esigenze poste dal problema da risolvere (Buchanan et al., 2017). La scelta di uno specifico modello/algoritmo di ML è tutta nelle mani degli esperti di ML e la sua adeguatezza dipende dalle loro specifiche competenze sia nell'ambito del ML che nel dominio di applicazione.

[12]*https://scikit-learn.org/stable/auto_examples/classification/plot_ classifier_comparison.html.*

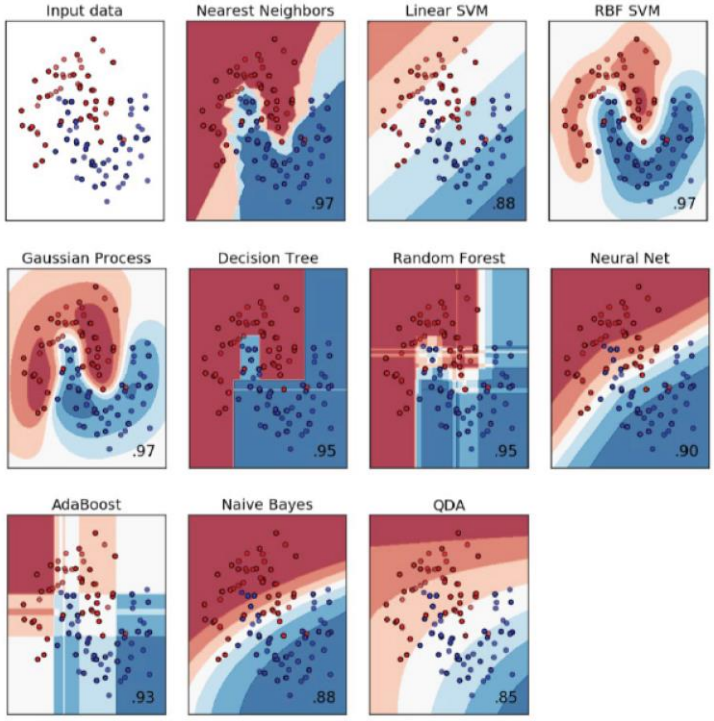


Fig. 2.5. *Classificazioni dello stesso insieme di dati di input di Figura 2.3(a) ottenute con modelli computazionali diversi.* Fonte: Pedregosa et al., 2011, figure tratte da sci-kit-learn "Classifier Comparison"[13]

Anche in questo caso, quindi, risulta difficile pensare che sistemi di IA costruiti secondo le conoscenze e le tecnologie all'attuale stato dell'arte possano essere utilizzati per l'esecuzione di funzioni militari critiche, come decisioni autonome – cioè senza alcun significativo controllo umano – relative alla vita e la morte di avversari, e questo non solo per ragioni di natura etica e di Diritto Umanitario Internazionale, come discusso nei Capitoli 1 e 7 del presente documento, ma anche sulla base di considerazioni tecniche.

Infine, va detto che nei sistemi di ML *supervised*, poiché, come si è visto, il *training* di un modello consiste nella calibrazione dei suoi parametri effettuata in maniera automatica, con iterazioni guidate da un'enorme quantità di dati, nella fase di uso della macchina risulta praticamente impossibile alla mente umana – incluse quelle degli stessi programmatori della macchina – comprendere perché, dato un certo input, essa produca uno specifico risultato: queste tecniche, quindi, producono sistemi di IA che vanno utilizzati come delle vere e proprie *black box*.

Questa caratteristica è particolarmente importante e critica per il fatto che, purtroppo, i sistemi di ML a volte, ed in maniera non facilmente predicibile, producono risultati errati e totalmente inaspettati. Per esempio, sono stati effettuati degli

[13]*https://scikit-learn.org/stable/auto_examples/classification/plot_ classifier_comparison.html*

esperimenti nell'area del riconoscimento automatico delle immagini in cui si è visto che leggere perturbazioni dei valori dei *pixel* dell'immagine di input, del tutto irrilevanti per gli umani, inducono la macchina a clamorosi errori, come quello di classificare l'immagine di uno scuolabus come quella di uno struzzo (Klarreich, 2016), (Edwards, 2019), (OpenAI, 2019); analogamente, il sistema di riconoscimento può fallire facilmente se si agisce direttamente sugli oggetti da riconoscere, nello spazio reale, del quale viene sottoposta al sistema un'immagine, per esempio una foto (Thys et al. 2019).

È evidente che comportamenti e situazioni come quelli ai quali si è accennato sopra non possono essere tollerati per (sotto-)sistemi impiegati per lo svolgimento di funzioni critiche, come quelle di selezione e decisione dell'abbattimento del bersaglio da parte di future armi autonome.

Per applicazioni meno critiche, il problema del superamento del limite della *black box* costituisce oggi una vera e propria sub-disciplina dell'IA che va sotto il nome di "spiegabilità" (*explainability*, in inglese) ed è attualmente oggetto di studio nel contesto di vari programmi di ricerca, come il programma *Explainable Artificial Intelligence (XAI)*, della *Defense Advanced Research Programs Agency (DARPA)* statunitense (Gunning et al. 2019), e i progetti *Local Interpretable Model-Agnostic Explanations* (Ribeiro et al. 2016) e *Human-Centered Artificial Intelligence* (HumanE AI, 2019). Trattandosi di ricerche ancora in corso riteniamo non opportuno tentare di esprimere valutazioni sull'effettiva applicabilità sul campo dei loro risultati.

Concludiamo questo paragrafo ritornando brevemente sul problema delle *vulnerabilità* dei sistemi di AI e ML ad attacchi informatici *specifici* per questa classe di sistemi – genericamente denominati "*adversarial attacks*" – e rimandando al capitolo 4 per una discussione più generale sui rischi di sicurezza informatica e *cyberwar*.

Un primo tipo di attacchi specifici per i sistemi di ML è quello dei cosiddetti *exploratory attacks,* tipici di avversari (anche umani) che si evolvono nel tempo, scoprendo, e quindi sfruttando, vari punti deboli degli specifici programmi di ML. Ad esempio, un sistema automatico *anti-spam* potrebbe fare uso di un vocabolario per identificare gli *spam*; uno spammer può quindi imparare a storpiare leggermente le parole, inserendo dei semplici "errori" ortografici e, in questo modo, evitare di essere scoperto. L'essenza, quindi, di questi attacchi risiede nell'indurre le macchine a considerare legittimi input che invece non lo sono. (Buchanan et al., 2017).

Un'altra classe di attacchi molto insidiosi è quella dei cosiddetti *causative attacks*. In questo caso, l'avversario cerca di creare delle debolezze nel sistema che sfrutterà in un secondo tempo. Un tipico attacco di questo tipo è quello del cosiddetto "avvelenamento" del *training data set*. Questo altro non è che un meccanismo per creare ad arte un *bias* nel sistema talmente forte che, in definitiva, lo porta a imparare "le cose sbagliate" e quindi poi commettere errori, anche importanti, nella classificazione effettuata durante il suo uso (Buchanan et al., 2017). Ad esempio,

avversari del nostro sistema *anti-spam* possono avvelenare i dati di *training* con una gran quantità di messaggi con contenuto pedo-pornografico e istruirlo in maniera da fare considerare questi messaggi legittimi; in questo modo, se l'*anti-spam* impara che la presenza di riferimenti alla pedo-pornografia sono indizio del fatto che un messaggio è "buono", non sarà in grado di bloccare gli *spam* di quel tipo. Questo tipo di attacchi sono particolarmente diffusi specie contro quei sistemi di ML che fanno *re-training*, cioè che continuano a (ri)calibrare i loro parametri anche sulla base dei dati che ricevono durante l'uso. Naturalmente, tutti i tipi di classificatori che basano la loro funzionalità sul *training* possono essere soggetti ad attacchi di tipo *poisoning*.

Inoltre, dobbiamo richiamare e sottolineare il fatto che i sistemi di IA, e in particolare di ML, stanno dimostrando capacità e *performance* estremamente interessanti quando applicati a domini molto specifici e per l'esecuzione di compiti ben delimitati. Questo tipo di IA viene normalmente classificata come IA "stretta" (*Narrow AI*, in inglese). Ad esempio, nel caso delle reti neurali, si assiste al cosiddetto "*catastrophic forgetting*": quando una rete cerca di apprendere nuove funzionalità o nuovi compiti, tipicamente dimentica quelli imparati in precedenza. Lo stato dell'arte in queste discipline non consente, al momento, di avere a disposizione sistemi di "IA generale", cioè sistemi realmente intelligenti, capaci, fra l'altro, di trasferire in altri domini conoscenze apprese in un certo dominio applicativo; attualmente, l'IA generale è solo un *obiettivo* della ricerca in IA. A maggior ragione, una "super IA", cioè sistemi che esibiscano un'intelligenza superiore a quella naturale/umana in tutti i domini della conoscenza non è realizzabile allo stato dell'arte né, si ritiene, nel prossimo futuro (UNIDIR, 2018).

A questo punto, risulta quanto meno problematico credere che "[i] [...] dati lavorano per te. Tu accumuli più dati possibile e li addestri a insegnare e addestrare sistemi autonomi [...]. [Lo] scopo del primo o del secondo giorno [di battaglia] non sarà [più] quello di uscire e distruggere aerei nemici o altri sistemi. Esso è [invece] quello di uscire, accumulare dati, fare ricognizione di dati, così che i nostri sistemi di apprendimento diventino più intelligenti di [quelli del nemico]"[14].

A nostro avviso, è più ragionevole tenere presente che "è possibile immaginare un attacco [...] di *poisoning* dei dati che potrebbe portare un sensore [di un sistema di ML] a classificare un amico come nemico o a non rilevare la presenza di un nemico" (Allen et al., 2017, traduzione in italiano a cura degli autori).

[14] Citazione di William Roper, capo dello *Strategic Capabilities Office* del Pentagono, secondo Patrick Tucker, 'The Next Big War Will Turn on AI, Says US Secret-Weapons Czar', Defense One, 28 March 2017, come riportato in (Buchanan et al. 17, p. 21-22); traduzione in italiano a cura degli autori.

2.4 Conclusioni

In questo capitolo abbiamo fornito una breve introduzione all'IA e alle tecniche di ML, e una discussione dei limiti e delle problematicità di queste ultime.

D. Parnas, (Parnas, 2017) sottolinea come sia opportuno e necessario sviluppare attività di ricerca in questo campo, con particolare riferimento alla *explainability*, e alla sicurezza riguardo agli attacchi tradizionali e *adversarial*. Inoltre, è importante che, prima di utilizzare questi sistemi fuori dai laboratori di ricerca, vengano valutate accuratamente e in maniera rigorosa tutte le implicazioni di natura etica e sociale, e, nel caso dell'utilizzo per scopi militari e di sicurezza internazionale, anche quelle di diritto internazionale.

Si sente spesso affermare, anche in ambienti scientifici, che di fronte a un problema del quale non si conosce la soluzione, piuttosto che affrontarlo con il metodo scientifico, partendo dall'osservazione e cercando di individuare cause ed effetti per sviluppare una teoria, sia più opportuno affidarsi all'IA, perché con sufficiente potenza di calcolo e sufficienti dati essa lo risolverà. Noi invece pensiamo che questo approccio sia sbagliato e pericoloso e che la costruzione di sistemi informatici affidabili debba comunque essere guidata da solidi principi di ingegneria dei sistemi e *trustworthy computing*, come sottolineato in (Cerf, 2019) e (Neumann, 2019).

Concludiamo quindi affermando che condividiamo appieno i punti di vista di eminenti personalità della comunità degli informatici, quali David L. Parnas, William G. Cerf e Peter G. Neumann, sulla ricerca in AI, e più in generale dell'informatica, e sul loro uso specie in campi di applicazione *critici*, dove si richiede il massimo possibile di affidabilità e, più in generale di *dependability*e *trustworthyness* (Avizienis et al. 2004). Allo stesso tempo, riteniamo che l'IA sia sicuramente una disciplina interessante e con utili applicazioni, specie se usata in specifici domini di applicazione e per compiti ben definiti.

Bibliografia

Aiello M., Pratt-Hartmann I., van Benthem J. (a cura di) (2007). *Handbook of Spatial Logics*. Berlin, D: Springer.

Allen G., Chan, T. (2017). “Artificial Intelligence and National Security”. In *STUDY 2017*. Cambridge (MA), USA: HARWARD Kennedy School. BELFER CENTER for Science and Int. Affairs.

Amoroso D., Sauer F., Sharkey N., Suchman L., Tamburrini, G. (2018). "Autonomy in Weapon Systems. The Military Application of Artificial Intelligence as a Litmus Test for Germany’s New Foreign and Security Policy". Berlin, D: The Heinrich Böll Foundation.

Avizienis A., Laprie J-C., Randell B., Landwehr C. (2004). "Basic Concepts and Taxonomy of Dependable and Secure Computing". In*IEEE Transactions on Dependable and Secure Computing,* 1(1):11-33.

van Benthem J., Blackburn P. (2006). "Modal Logic: A Semantic Perspective." In Blackburn P., van Benthem J., Wolter F. (a cura di), *Handbook of Modal Logic*. Amsterdam, NL: Elsevier.

Boulanin V., Verbruggen M. (2017). "Mapping the Development of Autonomy in Weapon Systems". Stocholm, SE: Stockholm International Peace Research Institute (SIPRI).

Buchanan B., Miller, T. (2017). “Machine Learning for Policymakers. What It Is and Why It Matters”. *PAPER 2017*. Cambridge (MA), USA: HARWARD Kennedy School. BELFER CENTER for Science and Int. Affairs.

Cerf V.G. (2019). “AI Is Not an Excuse!”. In *Communications of the ACM*, 62(10): 7.

CRS (2018). "U.S. Ground Forces Robotics and Autonomous Systems (RAS) and Artificial Intelligence (AI): Considerations for Congress". CRS Report, Washington, USA: Congressional Research Service.

Cummings M.L. (2017). "Artificial Intelligence and the Future of Warfare". The Royal Institute of International Affairs. London, UK: Chatham House.

Din A.M., (a cura di) (1986). “Arms and Artificial Intelligence. Weapon and Arms Control Applications of Artificial Intelligence”. Stocholm, SE: Stockholm International Peace Research Institute (SIPRI).

Dyndal G.L., Berntsen T.A., Redse-Johasen S. (2017). "Autonomous military drones: no longer science fiction". In *NATO Review Magazine,* 28 luglio 2017.

van Ditmarsch H., Halpern J. Y. (a cura di) (2015). *Handbook of Epistemic Logic*. London: UK. King's College Pubn.

DSB (2016). *Report of the Defense Science Board Summer Study on Autonomy*. Washington, USA: US Department of Defense - Defense Science Board.

Edwards C. (2019). “Hidden Messages Fool AI”. In *Communications of the ACM*, 62(1):13-14.

Emerson E.A., (1990). "Temporal and Modal Logic". In van Leeuwen J. (a cura di), *Handbook of Theoretical Computer Science*. Boston, USA: The MIT Press.

Franklin S. (2014). “History, motivation and core themes”. In Frankish K. e Ramsey W. M. (a cura di), *The Cambridge Handbook of Artificial Intelligence*. Cambridge, UK: Cambridge University Press.

GAO (2017). "Internet of Things. Status and implications of an increasingly connected world". Technology Assessment. Report to Congressional Requesters GAO-17-75. Washington, USA: United States Government Accountability Office (GAO). Center for Science, Technology, and Enginnering.

Goodfellow I., Bengio Y.Y, Courville A. (2016). *Deep Learning*. Cambridge (MA), USA: The MIT Press.

Gunning D., Aha D.W. (2019). "DARPA ExplainableArtificial Intelligence Program". In *AI Magazine*, 40(2):44-58.

HumanE AI (2019). *Human-Centered Artificial Intelligence*. EU Horizon2020 funded project https://www.humane-ai.eu (accesso effettuato il 17 ottobre 2019).

Klarreich E. (2016). “Learning Securely”. In *Communications of the ACM*, 59(11):12-14.

Landman N., Pang H., Williams C. (2019) "K-Means Clustering". *Brilliant.org.* Retrieved 18:44, October 13, 2019, from https://brilliant.org/wiki/k-means-clustering/

Neumann P.G. (2019). “How Might We Increase System Trustworthyness?”. In *Communications of the ACM*, 62(10):23-25.

Open AI (2019). “Attacking Machine Learning with Adversarial Examples” https://openai.com/blog/adversarial-example-research/(accesso effettuato il 17 ottobre 2019).

Parnas D.L. (2017). “The Real Risks of Artificial Intelligence”. In*Communications of the ACM*, 60(10):27-31.

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher

M., Perrot M., Duchesnay, E. (2011). "Machine Learning in Python". In *Journal of Machine Learning Research*, 12: 2825-2830.

Ribeiro M.T., Singh S., Guestrin C. (2016). ""Why Should I Trust You?": Explaining the Predictions of Any Classifier". InKrishnapuram B., Shah M., Smola A.J., Aggarwal C.C., Shen D., Rastogi R. (a cura di) *Proceedings of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, San Francisco, CA, USA, Aug. 13-17, 2016*. USA: Association for Computing Machinery (ACM)

Rossi J.C. (2019). "Un'opera dell'uomo: le macchine autonome letali". In *IRIAD Review*, 5, ArchivioDisarmo - IRIAD:2-22.

Thys S., van Ranst W. Goedeme' T. (2019). "Fooling automated surveillance cameras: adversarial patches to attack person detection". https://arxiv.org/abs/1904.08653 (accesso effettuato il 17 ottobre 2019).

UNIDIR (2018). "The Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence". UNIDIR Resources N. 8. Geneva: United Nations Institute for Disarmament Research (UNIDIR) - https://www.unidir.org/files/publications/pdfs/the-weaponization-of-increasingly-autonomous-technologies-artificial-intelligence-en-700.pdf

Wikipedia (2018). "Deep Learning" https://en.wikipedia.org/wiki/Deep learning (accesso effettuato il 9 giugno 2018).

Wikipedia (2019). "Tenso Processing Unit" https://en.wikipedia.org/wiki/Tensor_processing_unit (accesso effettuato il 9 giugno 2018).

Cap. 3
Applicazioni militari e i potenziali sviluppi dei LAWS

3.1 Introduzione

I recenti progressi nel campo della robotica militare, grazie anche all'era digitale, consentiranno in un futuro non troppo distante di realizzare robot con capacità sempre più indipendenti.

Vi sarà un'ennesima rivoluzione negli affari bellici in grado di imporre un cambio di paradigma nella guerra, rendendola più dinamica e letale di quella odierna. Armi completamente autonome ad oggi non sono ancora presenti, ma gli odierni sviluppi nel campo della robotica e della IA potrebbero a breve portare alla nascita dei *Killer Robot*, ovvero, macchine che se realizzate e configurate per rispondere ad eventuali condizioni, sarebbero in grado addirittura di ingaggiare un bersaglio senza la necessità di alcun input esterno.

Nel settembre 2017, il presidente Putin in un discorso alla TV nazionale ha dichiarato che "*l'intelligenza artificiale non è soltanto il futuro della Russia, ma è il futuro di tutta l'umanità…chi diventerà il leader in questo settore sarà il signore del mondo*".

Come risposta a settembre 2018, il Pentagono si è impegnato a compiere il più grande investimento fino ad oggi mai fatto in sistemi di intelligenza artificiale (IA) per gli armamenti statunitensi, impegnandosi a spendere 2 miliardi di dollari nei prossimi cinque anni attraverso la DARPA (*Defense Advanced Research Projects Agency*), per "*sviluppare la prossima ondata di tecnologie di IA*" (PAX, 2019c, p. 4).

È una chiara dimostrazione da parte delle potenze tecnologicamente più avanzate di come già da tempo sia iniziata una vera e propria corsa verso la dotazione di armamenti sempre più autonomi e letali.

Attori come gli Stati Uniti, la Russia, la Cina ed alcuni paesi della NATO, infatti, già da tempo hanno intrapreso questo nuovo percorso, apparentemente incuranti delle loro incognite e dei rischi legati ad un loro eventuale malfunzionamento; anzi, finanziano e sostengono svariati progetti, coinvolgendo così anche diversi settori tecnologici e relativi dipartimenti delle università.

Una prima chiave di lettura per comprendere come le maggiori potenze vogliano utilizzare queste innovazioni nei conflitti armati, passa proprio dalla spesa e dai programmi che questi destinano per il loro sviluppo e che in futuro non potrà che aumentare.

Gli **Stati Uniti**: il Paese, relativamente più trasparente in materia di difesa, si è sempre avvalso del *Department of Defence* (DOD) per stabilire i punti chiave della sua

politica in termini di finanziamenti e di priorità nella Ricerca e Sviluppo (R&S) della IA.

Una ulteriore conferma di quella sua volontà mai troppo celata di voler soverchiare la concorrenza in questo ambito può essere intravista in quell'*offset strategy*, del 2014, la terza nella storia americana (Petroni, 2015, p. 143)[15].

L'obiettivo di questa strategia, secondo Bob Work, all'epoca segretario alla difesa, è quello di sfruttare tutti i progressi dell'intelligenza artificiale e dell'autonomia per poi inserirli nei network di battaglia del DoD, ottenendo così un aumento delle prestazioni che il dipartimento ritiene rafforzerà la propria deterrenza (The White House, 2016).

Nel 2015, poi, il DOD ha speso ben 149 milioni di dollari ed un totale di 18 miliardi sono stati destinati a investimenti continui per il periodo 2016-2020. Le quote, sono state così distribuite: alla DARPA il 29%, all' *Office Of Naval Research* il 25%, all'*Air Force Research Laboratory* il5% ed il restante 28% destinato alla R&S all'interno di università e società private (Boulanin, Verbruggen, 2017, p. 94). La maggior parte dei progetti della DARPA in R&S, pur se molto ambiziosi, ha finalità meramente esplorativi.

Altri programmi altrettanto ambiziosi ed in fase di sviluppo sono:

- il *Collaborative Operations in Denied Enviroment* (Code)[16]: tra i più rilevanti in fase di sviluppo, mira a sviluppare nuovi algoritmi o software per gli *Unmanned Aerial Vehicle* UAV già esistenti che estenderebbero la loro capacità di missione e migliorerebbe la capacità delle forze statunitensi per condurre un attacco in uno spazio aereo ostile pur se sempre sotto il controllo di un operatore. Inoltre, i velivoli dotati di tale sistema sarebbero in grado non solo di scovare obiettivi e di impegnarli secondo le regole di ingaggio stabilite, ma anche di adattarsi a situazioni dinamiche come l'emergere di minacce impreviste. Diversi test sono stati già effettuati in quest'ambito da compagnie quali Lockheed Martin e Raytheon (PAX, 2019c, p. 9).
- il programma ATLAS[17]: un sistema automatizzato avanzato di *targeting*, il quale prevede l'utilizzo della IA e dell'apprendimento automatico per fornire ai veicoli di terra capacità di bersagli autonomi che consentirà alle armi di acquisire, identificare e ingaggiare un obiettivo in tempi almeno tre volte più rapidi dell'attuale processo manuale.

In aggiunta, a marzo del 2019, un progetto del Pentagono potrebbe portare alla costruzione della prima nave da guerra senza pilota prodotta in larga scala. Il progetto *Overlord* di SCO, arrivato ad ottobre alla sua seconda fase, ha come obiettivo quello di sviluppare l'autonomia al fine di realizzare e schierare in campo prototipi navali di superficie senza pilota come gli *Unmanned Surface Vehicles* (USV) in grado di funzionare perfettamente con la flotta[18].

---

[15] La prima, imperniata sulle armi nucleari negli anni 50, la seconda, invece, sul bombardamento di precisione.

[16]https://www.darpa.mil/program/collaborative-operations-in-denied-environment

[17]https://breakingdefense.com/2019/03/atlas-killer-robot-no-virtual-crew man-yes/.

[18]https://www.defensenews.com/naval/2019/03/14/a-classified- pentagon-maritime-drone-program-

Infine, gli Stati Uniti sponsorizzano progetti anche al di fuori del DOD. Tra questi, si segnala la *National Robotic Initiative*, che ha finanziato progetti di robotica con varie agenzie governative dal 2011. Con un budget di 221 milioni di dollari, nel 2017, di cui 103 milioni di dollari sono stati destinati al DOD ed il restante alla ricerca civile (Boulanin, Verbruggen, 2017, p. 97).

Il **Regno Unito**: è tra i membri della NATO che più investono in questo settore. A condurre la ricerca in questo ambito sono soprattutto Qinetiq[19] ed il DSTL (*Defence Science and Technology Laboratory*) agenzia esecutiva del Ministero della Difesa. Mentre Qinetiq svolge ricerche soprattutto in collaborazione con aziende private ed università. Il DSTL non solo destina il 40% del suo badget interno per quest'area, ma inoltre, finanzia con il 60% del suo budget anche la ricerca in aziende ed università. Dal 2010, poi, finanzia una serie di progetti di R&S per un totale stimato di circa 16 milioni di sterline (21 milioni di dollari) (Boulanin, Verbruggen, 2017, p. 97).

Nel dicembre del 2018, un rapporto del Ministero della Difesa ha promosso un maggiore uso della IA per respingere eventuali ritorsioni militari, sottolineando in aggiunta, il perseguimento della sua modernizzazione (PAX, 2019c, p. 19). Tra i vari sforzi del Ministero, va segnalato, inoltre, il programma *Autonomy* che basa le sue attività di ricerca su quelle tecnologie volte ad accrescere le capacità belliche del Paese. Proprio al fine di implementare il programma *Autonomy*, è sorto il *Defence Capability for Autonomous and Novel Technologies* (DECANT) ovvero un *framework* per costituire una rete di fornitori per le tecnologie dei sistemi autonomi[20].

Ancor più indicativo dell'interesse di Londra nei confronti delle nuove tecnologie è la cooperazione tra settore pubblico e privato già da tempo in atto. Le iniziative intraprese, infatti, pur partendo con lo sviluppo di semplici applicazioni di sicurezza, potrebbero costituire un eventuale trampolino di lancio per un più ampio uso militare. Un esempio in questo caso è l'hackathon[21] di swarm di droni del marzo 2019, organizzato dal Dstl e l'*Us Air Force Research Laboratory* (AFRL), che, pur se mira a sviluppare nuovi utilizzi di UAS per il soccorso e l'assistenza nella gestione di incendi, ha come obiettivo finale il miglioramento di algoritmi di IA e del *machine learning*. Inoltre, il Ministero della Difesa al suo interno può contare anche sul *Defence and Security Accelerator* (DASA), una organizzazione intergovernativa che avviato nel 2016 ha come compito quello di trovare e finanziare quelle innovazioni in grado di supportare la sicurezza e la difesa del Paese, al fine di mantenere così un vantaggio strategico nei

is-about-to-get-its-moment-in-the-sun/ - https://www.navsea.navy.mil/Media/News/Article/1976701/new-phase-of-ghost-fleet-overlord-unmanned-surface-vessel-testing-begins/

[19]ex agenzia di ricerca della difesa, privatizzata dal 2001.

[20] https://www.gov.uk/guidance/dstl-decant-framework.

[21]Binomio trahacker e thon (maratona), col quale si intende un evento destinato a informatici e dedicato alla collaborazione intensiva su un progetto comune.

confronti della concorrenza.

Nel 2019, infatti, con quasi 3 milioni di euro[22], nell'ambito del progetto *Many Drones Make light Work*, il Dasa ha assegnato il suo più grande contratto sino ad oggi. Il direttore di Blue Bear Systems commentando l'aggiudicazione del contratto ha dichiarato: "la capacità di schierare lo swarm in sistemi autonomi a basso costo offre un nuovo paradigma per le operazioni sul campo di battaglia"[23].

La Francia: pur di non perdere il passo con i suoi partner, Regno Unito e Stati Uniti, sta finanziando tramite il governo la ricerca accademica e industriale per le applicazioni militari. Nel gennaio 2017 il ministero della ricerca e dell'istruzione superiore ha pubblicato un rapporto sulla IA il quale ha formulato una serie di raccomandazioni come la creazione di *initIA*, fondo di ricerca che con 550 milioni di euro ha come scopo quello di fermare la fuga dei cervelli dei principali esperti nel settore (Boulanin, Verbruggen, 2017, p. 98).

Con la pubblicazione del Rapporto Villani, nel marzo del 2018, il Paese ha inoltre trovato una propria strategia nazionale sull'intelligenza artificiale. Il Rapporto si concentra su quattro settori prioritari per lo sviluppo della IA: sanità, mobilità e trasporti, ambiente, difesa e sicurezza[24].

Da qui la necessità di costruire sinergie attorno all'innovazione civile e militare nell'ambito dell'intelligenza artificiale. Come conseguenza, il Presidente Macron ha poi annunciato che il governo avrebbe investito un miliardo e mezzo fino al 2022. Investimenti che, facenti parte di una strategia più ampia, mirano a implementare quei sistemi di IA, che fanno uso del *Deep Learning*[25].

La Corea del Sud: settima al mondo per consistenza delle forze armate, la Repubblica di Corea non ha trascurato nemmeno il settore della robotica e grazie infatti a numerosi investimenti, Seoul si è affermato come leader del mercato nel campo dei robot industriali. Nel 2016, il governo sudcoreano ha inoltre deciso di investire 840 milioni di dollari tra il 2016 e il 2020 per incrementare la ricerca e lo sviluppo nell'intelligenza artificiale. Inoltre, il governo finanzierà l'istituzione della ricerca a guida civile di sei conglomerati (*chaebol*), che insieme investiranno 2,6 milioni di dollari (Boulanin, Verbruggen, 2017, p. 101).

L'esercito sudcoreano, all'inizio del 2019, ha avviato un centro di ricerca e di sviluppo sull'intelligenza artificiale. L'obiettivo è quello di costruire la visione e il

[22] https://www.gov.uk/government/news/25m-injection-for-drone-swarms.
[23] Ibidem.
[24] https://www.ilsole24ore.com/art/macron-punta-sull-intelligenza-artificiale-un-miliardo-e-mezzo-investimenti-AE2Mf3PE?refresh_ce=1
[25] https://www.politico.eu/article/macron-france-to-invest-nearly-e1-5-billion-for-ai-until-2022/

concetto per le applicazioni militari dell'IA e di sviluppare la prossima generazione di armi da combattimento.

Allo stesso scopo, nel febbraio 2018 il *Korea Advanced Institute of Science and Technology* (KAIST) ha annunciato che avrebbe collaborato con *Hanwha Systems*[26]. Proprio questa cooperazione ha istituito poi il Centro di ricerca per la convergenza della difesa nazionale e dell'intelligenza artificiale (Saalman, 2019, p. 34), il quale conduce ricerche relative sia alla guerra informatica sia alla robotica IA.

Uno dei programmi più importanti del Ministero della Difesa Nazionale sudcoreano (MND) riguarda gli *Unmanned Aerial Vehicles* (UAVs) identificati come uno dei nuovi settori per il rafforzamento del suo apparato militare. Già nel 2017, infatti, l'esercito sudcoreano aveva annunciato l'intenzione di creare un'unità specializzata, denominata *Dronebot Jeontudan*per la gestione degli UAV e veicoli terrestri senza pilota (UGV) (PAX, 2019c, p. 31).

Tre sono le motivazioni che spingono il Paese a dotarsi di questi mezzi: In primo luogo, aggirerebbe le difficoltà dovute al carattere montuoso della penisola fornendo, inoltre, una migliore sorveglianza. In secondo luogo, i suoi vicini, compresa la Corea del Nord, hanno già utilizzato sistemi autonomi e droni *swarm*. Nel 2014 e 2016, infatti, UAV nordcoreani sono stati individuati mentre si intromettevano nello spazio aereo sudcoreano. In terzo luogo, gli UAV possono essere utilizzati per contrastare le provocazioni non nucleari della Corea del Nord senza il costo di vittime umane (Saalman, 2019, p. 34).

La Russia: la nuova dottrina militare russa, pur se con un carattere non aggressivo, ha come obiettivo quello di contrastare l'egemonia statunitense. Ovvio quindi che Mosca riponga molta attenzione verso le armi belliche di nuova generazione. In realtà al Cremlino manca un vero programma nazionale per lo sviluppo della IA, lacuna che ha iniziato a colmare solo nel febbraio del 2019, allorquando per ordine di Putin ne sono iniziati i lavori. Questo però non deve trarre in inganno, in quanto la robotica militare svolge ormai da tempo un ruolo prioritario nella nuova politica di riarmo della Russia, con un budget stimato di circa di 346 miliardi di dollari per il periodo 2016-2025 (Boulanin, Verbruggen, 2017, p. 98).

Sviluppi significativi, in questo contesto, sono stati il lancio di programmi quali: il Robotics 2025 e la creazione dello *Skolkovo Robotics Center* (SRC) nel 2014. Il primo, il Robotics 2025, è implementato dal Ministero della difesa e dalla *Foundation for Advanced Studies* (FPI), nuovo centro di ricerca ed equivalente russa del DARPA. Il budget totale del FPI nel 2015-2016 è stato di 78 milioni di dollari di cui una parte è stata destinata per la ricerca e lo sviluppo della robotica (veicoli senza pilota, robot biologici, micro e nano) (Boulanin, Verbruggen, 2017, p. 98). Il secondo programma,

[26]la divisione della Difesa del Gruppo Hanwha, per la ricerca e lo sviluppo di armi di intelligenza artificiale.

l'SRC, invece, nasce con l'intento di migliorare la sinergia tra 35 istituti di ricerca: statali, universitari e aziendali, che insieme rappresentano il 20-25% tutte le entità russe coinvolte nel campo della robotica civile. Inoltre, SRC collabora sia con FPI sia con altri istituti di ricerca di altri Paesi, in particolare quelli cinesi.

Progressi significativi, sono invece quelli ottenuti da Kalashnikov, il più grande produttore di armi della Russia che nel 2017 ha annunciato di aver sviluppato un modulo di combattimento completamente automatizzato basato su tecnologie di *reti neurali* che, se confermato, consentirebbero sia di identificare obiettivi sia di prendere decisioni. La compagnia ha dichiarato che in futuro presenterà una serie di modelli basati su questa tecnologia: infatti, ad oggi un modulo di combattimento che sfrutta una tale tecnologia non è ancora presente sul mercato.

Tutti gli sforzi fin qui mostrati, complice anche quel suo timore per la sicurezza, dimostrano come anche la Russia sia ormai catapultata in quella corsa volta ad ottenere il dominio in questo campo.

La Cina: Ormai pronta ad entrare in un'*era d'oro* per lo sviluppo della IA, la Cina, attribuisce infatti alla IA un'alta priorità per accrescere il proprio prestigio internazionale.

Pechino, in realtà ha iniziato a muovere i primi passi in questo settore già dal 1986, con il programma 863 (piano di sviluppo di alto livello statale) dove veniva stabilita la base delle attività di R&S per la IA e la robotica, con l'obiettivo di sviluppare un'innovazione nei settori tecnologici più avanzati, in particolare l'informatica e l'automazione.

Nel 2014 poi, la Cina ha ufficialmente sottolineato come la robotica e la IA siano aree tecnologiche prioritarie. Da allora, infatti, una serie di piani economici e scientifici ha preso il via, tra cui il 13° piano quinquennale per lo sviluppo economico e sociale della Repubblica Popolare Cinese (2016-2020).

Nel 2016 Pechino ha poi pubblicato due piani dedicati alla IA e alla robotica: il piano dello sviluppo dell'industria della robotica (2015-2016) e la guida triennale per internet plus (2016-2018). Quest'ultimo ha poi aperto la strada a tre centri di ricerca congiunti, guidati da *Baidu*, l'equivalente cinese di Google per il *deep learning*, i *big data* e la IA.

Infine, Il *Piano di sviluppo dell'intelligenza artificiale di nuova generazione* del 2017 delinea la strategia cinese in tre mosse: la prima fase è quella di mantenere la tecnologia globale e l'applicazione della IA con il livello avanzato del mondo entro il 2020. La seconda, invece, è quella di ottenere rilevanti scoperte in termini di teoria della IA base entro il 2025. Ed infine, la terza, è che entro il 2030 il Paese si affermi come leader mondiale in questo settore. Tutti questi passaggi, permetteranno di utilizzare la IA per la sicurezza e il controllo della difesa nazionale.

A preoccupare i suoi avversari non è solo il rapido sviluppo tecnologico di Pechino, ma anche quella sua assenza di etica militare per limitare l'utilizzo di queste nuove

tecnologie. Secondo il Pentagono, infatti, a breve si prevede che presto la Cina potrebbe iniziare ad esportare anche ai Paesi del Medio Oriente velivoli di nuova generazione[27].

3.2 Stato dell'arte

Le grandi potenze possono già vantare a disposizione nei loro arsenale armi semi-autonome di vario tipo quali: gli *Unmanned Combat Aerial Vehicle* (UCAV), gli *Unmanned Underground Vehicles* (UGV), i *Close in Weapon System* (CIWS)etc. tutti in grado di svolgere operazioni in semi-autonomia pur se sotto la supervisione umana. Tuttavia, altro non costituiscono che un preludio ai cosiddetti *Killer Robot*. Proprio per le capacità che questi sistemi dimostrano, sia quelli già schierati in campo sia quelli in progettazione, appare opportuno uno specifico approfondimento in quanto mostrano un primo quadro del grado di avanzamento della tecnologia e della corsa agli armamenti intrapresa dai vari Stati.

3.3 Munizioni Loitering

Pur mancando una definizione precisa, con tale terminologia si fa riferimento a quelle *munizioni da bombardamento* o *droni kamikaze*, ovvero, una sorta di ibrido tra un drone e un missile guidato. Le munizioni *Loitering* sono un tipo di velivoli senza equipaggio armati e progettati non per colpire un bersaglio predefinito, ma una determinata area con una testata esplosiva. Disponibili in varie dimensioni[28] sono, inoltre, dotate di telecamere elettro-ottiche e ad infrarossi che consentono così di localizzare, sorvegliare e guidare il velivolo verso il bersaglio (Boulanin, Verbruggen, 2017, p. 53). Caratteristica operativa che le contraddistingue è la capacità di *svolazzare* in aria per un tempo prolungato dando così la possibilità di trovare, decidere e colpire un bersaglio sul campo di battaglia. Sono utilizzabili sia in missioni offensive sia in missioni difensive che potrebbero essere ritenute pericolose o rischiose per altri tipi di sistemi senza equipaggio. Alcuni modelli, inoltre, come quelli utilizzati per la SEAD (*Suppression of Enemy Air Defences*), possono funzionare addirittura in completa autonomia dopo il lancio.

Un problema che si sta sempre più riscontrando è che, a differenza degli UCAV o

[27] https://www.9news.com.au/technology/killer-drones-sold-by-china-to-middle-east-warns-us-defence-secretary-mark-esper-military-technology-news/6acddc00-20b9-4316-99d4-71f5aca76be5?ref=BP_RSS_ninenews_0_china-selling--killer-robots--to-middle-east_071119

[28] Alcuni, quelli più piccoli, possono essere trasportati in uno zaino dalle truppe, mentre quelli più grandi hanno una dimensione pari a quella dei missili con un peso fino a 32 kg. I modelli più grandi, inoltre, hanno potenzialmente la capacità di tornare alla base se non trovano alcun obiettivo o se la missione viene interrotta. Tuttavia, la maggior parte di questi si autodistrugge in volo.

degli *Unmanned Underwater Vehicles* UUV, il loro basso costo, consente ad un sempre più vasto gruppo di Paesi (e attori non statali) di possedere tali armi. A primeggiare nel settore sono gli Stati Uniti che, con Israele, hanno aperto la strada allo sviluppo di questo genere di armi.

MINI HARPY: la nuova munizione prodotta dalla IAI (*Israel Aerospace Industries*), combina le capacità dei modelli precedenti Harpy e Harop. Pensato per fornire agli operatori un controllo fino all'ultimo minuto, incluso la possibilità di cessazione dell'attacco in qualunque fase. Secondo Boaz Lezy, direttore generale e vice presidente esecutivo della IAI: "*anziché affidarsi a un punto di riferimento preciso, il sistema…si aggira in aria aspettando che il bersaglio compaia per poi attaccare e distruggere la minaccia*" (PAX, 2019b, p.22).

Caratteristiche tecniche	
Paese	Israele
Prodotto da:	Israel Aerospace Industries (IAI)
Ruolo:	Munizione loitering tattica multiuso
peso:	40Kg
Velocità:	55-85 nodi
Velocità massima:	200 nodi
Testata:	8kg
Autonomia di:	100 chilometri
Durata:	120 minuti
Venduto a:	in sviluppo

Tab. 3.1. *Mini Harpy.* Fonte: IAI, 2019

ORBITER 1K MUAS: progettato per colpire ed attaccare sia le persone sia veicoli leggeri con una testata a frammentazione esplosiva con palline di tungsteno. Pur se con un modus *man in the loop*, una volta lanciato può scovare, tracciare e attaccare bersagli in completa autonomia. Nel 2017, il Ministero della Difesa israeliano ha bloccato la vendita dell'Orbiter in Azerbaigian, in quanto la dimostrazione lì condotta lo impegnava contro obiettivi armeni (PAX, 2019b, p.22).

Caratteristiche tecniche	
Paese:	Israele
Prodotto da:	Aeronautics Ltd.[29]
Durata:	2.5 ore
Peso massimo:	13 kg
Velocità:	30/40 nodi
Mission range:	maggiore di 100 km

Tab. 3.2. *ORBITER 1k MUAS.* Fonte: Aeronautics, 2019

KARGU: STM, società di proprietà statale turca, sta attualmente implementando le capacità del suo bombardamento tramite l'utilizzo della IA, incluso il riconoscimento facciale. Con una capacità di *fire and forget*in autonomia attraverso l'inserimento delle coordinate del bersaglio, è un UAV ad ala rotante, ideato per scenari di guerre anti-terrorismo e asimmetriche in quanto può rispondere in maniera rapida contro obiettivi fissi e mobili. È inoltre possibile far funzionare 30 Kargu in *sciame* permettendo così di distruggere addirittura un'unità militare o una nave da guerra. Secondo fonti ufficiali, i Kargu saranno schierati nella parte orientale della Turchia ai confini con la Siria[30].

Caratteristiche Tecniche	
Paese:	Turchia
Prodotto da:	STM
Peso:	7kg ciascuno
Autonomia:	15 chilometri
Tempo in volo	30 minuti
Venduto a:	Turchia

Tab. 3.3. *Kargu.* Fonte: STM, 2019

SWITCHBLADE: ampiamente utilizzato dalle truppe americane in Afghanistan, è stato rilevato anche in Iraq e Siria. Progettato per la sorveglianza e la ricognizione (ISR) su un obiettivo BLOS (al di là del campo visivo) in pochi minuti. Inoltre, «*garantisce una soluzione di attacco con un'elevata precisione e con effetti collaterali minimi*»[31].

È una munizione *loitering* lunga 0,6 m con una testata del carico utile modulare di 40 mm. Sono stati osservati diversi modelli e «*può essere azionato manualmente o autonomamente*» (PAX, 2017, p. 9).

[29] Società acquistata da Rafael nel 2019.

[30] https://www.newscientist.com/article/2217171-autonomous-killer-drones-set-to-be-used-by-turkey-in-syria/

[31]https://www.avinc.com/images/uploads/product_docs/SB_Datasheet_2017_Web_rv1.1.pdf.

Caratteristiche Tecniche	
Paese:	USA
Prodotto da:	Aero Vironment (USA), Orbital ATK
Ruolo:	surveillance and reconnaissance ISR - beyond-line-of-sight (BLOS) targets.
Peso:	2,5 kilogrammi
Range:	10 kilometri, con un backpack fino a 15 - 45 km
Tempo di volo:	10 minuti circa
Venduto a:	US Army

Tab. 3.4. *Switchblade.* Fonte: AeroVironment – AV, 2019

WARMATE 2: versione più grande della precedente, il lancio avviene tramite una catapulta con una testata anticarro, anti implosione (termobarica). La varietà dei diversi effetti ottenuti varia a seconda della testata utilizzata. A quest'ultimo modello, è stata aggiunta una termocamera (IR infrarossi) nella fusoliera, consentendo così non solo il target automatico, ma anche la possibilità di valutare un obiettivo prima di colpirlo.

Caratteristiche Tecniche	
Paese:	Polonia
Prodotto da:	WB Electronics
Tempo di volo:	120 minuti
Velocità massima:	150 km/h
Telecamere	EO e IR – opzionali, inseriti nella fusoliera

Tab. 3.5. *Warmate 2.* Fonte: WBGroup, 2019

Altri UAVs altrettanto rilevanti e sempre di società Israeliane sono: lo **Skystriker** di Elbit systems, un *Unmanned Aerial System* (UAS) che dispone di un sistema di navigazione autonoma durante le diverse fasi e per di più, può essere bloccato su un bersaglio dall'operatore mediante un sensore elettro-ottico. UVision, invece, presenta la sua famiglia di munizioni *loitering* **Hero** con a disposizione diverse munizioni e configurazioni. Secondo la compagnia, inoltre, ha la capacità di interrompere le operazioni fino all'ultimo secondo. Mentre per la Cina sono presenti i modelli: **ASN-301** di AVIC, definito come un sistema mobile anti-radiazioni e mostrato per la prima volta ad Abu Dhabi nel 2017 il quale presenta molte similitudine con il modello israeliano Harpy[32], ed i modelli **CH-901,** una mitragliatrice esagonale, prodotta da ALIT, consociata della CASC. Sempre della CASC vi è poi il modello **WS-43**. Infine, la Zala Aero[33], unica compagnia a sviluppare UAVs per la Russia, ha presentato i suoi modelli *loitering* **KYB** e

[32]Israele ha venduto il "drone suicida" Harpy alla Cina negli anni '90. Da: https://www.israeldefense.co.il/en/node/28716.
[33]Sussidaria del gruppo Kalashnikov.

Lantset nella prima parte del 2019 (PAX, 2019b, p. 23-24).

3.4 I sistemi d'arma ravvicinata o a corto raggio (CIWS)

Questa tipologia di arsenale per la difesa esiste già da molto tempo: infatti, il Mark 56, primo del suo genere, è stato ideato durante la Seconda guerra mondiale. Ben 89 Paesi dispongono di tale tipologia di armi e ben 63 di questi ne hanno schierate, col tempo, diverse varianti. Tra i maggiori produttori figurano gli Stati Uniti e la Russia (Boulanin, Verbruggen., 2017, p.37).

Il loro obiettivo in tutti questi casi è quello di rilevare, tracciare, selezionare la priorità ed attaccare la minaccia in modo più rapido e preciso di quanto potrebbe fare un operatore umano. Tuttavia, pur se la loro tecnologia è progredita, il loro funzionamento rimane quello del Mark 56.

Con riferimento ai CIWS volti alla difesa missilistica, invece, possiamo distinguerli in tre categorie: 1) in base alla gamma di utilizzo, come i GoolKeeper o i Phalanx, studiati per difendere una zona limitata; 2) i sistemi di difesa missilistica come l'Iron Dome, invece, possono fornire protezione su un ampia area geografica; 3) in base agli obiettivi che possono intercettare come missili razzi o aerei nemici, come il Centurion C-RAM, un CIWS terrestre ideato contro proiettili in arrivo, mentre il Phalanx (che opera su navi) può difendersi anche da veicoli di superficie.

La maggior parte di questi, infatti, può adottare come tipo di contromisura un *Hard Kill*, per sconfiggere le minacce in arrivo: tuttavia, alcuni modelli possono utilizzare addirittura un *Soft Kill*[34].

Tuttavia, i sistemi difensivi di seguito riportati possono essere classificati come automatici, in quanto, sono *fermi* o *fissi* su navi o rimorchi e progettati per sparare a bersagli inanimati. Infatti, eseguono ripetutamente azioni pre-programmate e operano solo all'interno di parametri e intervalli di tempo strettamente impostati in ambienti strutturati e controllati in modo comparabile[35].

Phalanx (MK15): a fuoco rapido della US Navy, progettati per avvertire dell'eventuale avvicinamento di missili antinavi, si presume che siano stati sperimentati anche in forma offensiva dagli Stati Uniti già nel 1980[36]. Esiste anche una loro versione terrestre, i C-RAM, che esperti militari ritengono siano stati, invece, impiegati in Iraq

[34] Con il termine *Hard Kill* si intendono tutte quelle misure che fisicamente contrattaccano una minaccia in arrivo, mente, con Soft Kill quelle contromisure elettroniche che modificano la firma elettromagnetica o di altro tipo del sistema mirato, alterando così il rilevamento della minaccia in arrivo (es. missile guidato).

[35] https://www.armscontrol.org/act/2016-09/features/stopping-%E2% 80%98killer-robots%E2%80%99-why-now-time-ban-autonomous-weapons-systems

[36] Chansoria M., Roff H., Santos P. E., *Autonomous weapons, civilian safety, and regulation versusprohibition*, *Autonomous weapons*: *tightrope balance*, 2016. Consultabile in http://thebulletin.org/ autonomous-weapons-civilian-safety-and-regulation-versus-prohibition

nel 2005[37].

Caratteristiche Tecniche	
Paese:	Stati Uniti
Prodotto da:	Raytheon Systems Company (preceduto da Hughes Missile Systems Company e acquistato da General Dynamics Pomona Division nel 1992)
Ruolo:	Anti-ship missile defense
Peso:	6,120 kilogrammi
Calibro:	20mm
Venduto a:	US Navy

Tab. 3.6. *Phalanx.* Fonte: America's NAVY, 2019

Centurion C-RAM: Il *Counter-Rocket, Artillery, Mortar*, sviluppato in parte da NorthropGrumman, è stato inizialmente pensato per proteggere le basi statunitensi in Iraq[38]. Il C-RAM, in realtà, altro non è che la versione terrestre del Phalanx e per questo presenta funzioni simili. Tra i due modelli, tuttavia, sussistono alcune differenze significative, in quanto i proiettili del CIWS rischiano di finire in oceano se mancano il bersaglio. Sulla terra è molto più probabile che causino, invece, danni collaterali. Per questo a differenza della sua controparte quelli del C-RAM si autodistruggono prima che colpiscano il suolo.

Caratteristiche tecniche	
Paese:	Stati Uniti
Prodotto da:	Northrop Grumman (primo contraente), Raytheon, L3 Communications, Lockheed Martin
Munizioni:	20 mm HEIT-SD (High-Explosive Incendiary Tracer, Self-Destruct)
Range:	1,2 km
Tasso di abbattimento:	60 – 70%

Tab. 3.7. *Centurion C-RAM.* Fonte: GlobalSecurity, 2018

Iron Dome: Dalla corsa verso la dotazione di armamenti autonomi non si è certo ritirato Israele che, vivendo sin dalla sua nascita situazioni di insicurezza territoriale, con i suoi Iron Dome[39], dotati di venti missili intercettori Tamir, già da tempo difende i suoi confini con Gaza e protegge i suoi interessi geostrategici nella Penisola del Sinai

[37] http://www.navweaps.com/index_tech/tech-103.php

[38] https://www.globalsecurity.org/military/systems/ground/cram.htm

[39] È progettato per individuare missili a corto e medio raggio e proiettili d'artiglieria da 155mm fino a 70 chilometri di distanza, grazie all'ausilio di un radar. Il sistema è capace, inoltre, di intercettare attacchi con un successo maggiore dell'80%. Cfr. HRW, *Losing of Humanity,* cit., pp. 10-Consultabile in: https://www.hrw.org/sites/default/files/reports/arms1112ForUpload_0_0.pdf.

(Rossi, 2016). Un sistema di difesa missilistico che si basa su tre componenti essenziali: un radar EL/m-2081 in grado di individuare i razzi e tracciarne la traiettoria; un sistema di gestione BMC capace di determinare dove atterrerà l'ordigno, valutarne così la minaccia e calcolare il punto d'impatto in aria; l'unità di lancio dei missili, dispositivo mobile, dotato di 20 missili Tamir, dotati di sensori elettro-ottici, guida GPS e otto pinne per virare in aria. Per ogni minaccia da intercettare, vengono lanciati 2 missili Tamir, uno per aprire la zona d'impatto e un altro per distruggere la minaccia.

Caratteristiche tecniche	
Paese:	Israele
Prodotto da:	Rafael Advanced Defense Systems
Ruolo:	Difesa antiaerea
Peso:	198 lb (90 kg) missile
Capacità:	20 missili per batteria
Gittata:	70 km circa
Radar:	EL/m-2081 della ELTA
In serivzio:	12 batterie, 15 come obiettivo finale
Costo:	50 millioni di dollari per batteria
Venduto a:	Israele

Tab. 3.8. *Iron Dome.* Fonte: ArmyTechnology, 2019

NBS Mantis: già operativi sul campo si possono ritenere appartenenti a questa categoria, anche, gli NBS Mantis che, schierati dalla Germania in Afghanistan, proteggono con la loro presenza le proprie basi in quel Paese (HRW, 2012, p.10). Sviluppato da Rheinmetall Air Defence, per la difesa aerea, è dotato di un sistema di controllo a terra, due sensori e 6 Oerlikon 35 mm altamente automatizzati "*Milennium Guns*"[40].

Caratteristiche Tecniche	
Paese	Germania
Produttore	Rheinmetall
Peso	5800 kg
sensori	2 sensori elettro-ottici
Armi:	6 Oerlikon altamente automatizzati da 35 mm " Millennium guns"

Tab. 3.9. *Mantis.* Fonte: ArmyTechnology, 2019

GoalKeeper SGE30: progetto congiunto tra la Signaal e la General Electric ed ora di proprietà di Thales. Questo CIWS è un'arma ravvicinata contro missili e aerei. Il sistema

[40] http://www.defenseindustrydaily.com/Germany-Orders-Skyshield-C-RAM-Base-Defense-Systems-05418

esegue automaticamente sia il processo dalla sorveglianza sia quello della rilevazione. Può tracciare fino a trenta obiettivi e selezionare poi tra i quattro più pericolosi, per poi sparare per circa 0,2 secondi per ingaggio. Introdotto inizialmente alla fine degli anni 70, ha avuto diversi upgrade. Thales, infatti, sta migliorando le capacità optroniche del sistema, i suoi algoritmi ed, inoltre, sta aggiungendo una nuova console per l'operatore.

Caratteristiche Tecniche	
Paese	Paesi Bassi
Prodotto	Signaal e General Electric prima, e poi da Thales
Peso:	6.730 kg
Cadenza di tiro:	70 copli/sec
Gittata massima:	più di 3000m
Venduto a:	numerosi Paesi in tutto il mondo, inclusi i Paesi Bassi, il Portogallo e la ROK (Rep. Di Corea)

Tab. 3.10. *Goalkeeper SG30.* Fonte: NavWeaps, 2017

3.5 Aerei da combattimento senza equipaggio UAV

Da sempre gli aerei da combattimento costituiscono per i governi un metodo per rassicurare/scoraggiare e sancire la propria forza militare. Con gli attuali progressi in questo campo, c'è una pressione su tutte le forze militari aeree per sviluppare ed implementare gli UCAV. Ovvero, aerei da combattimento senza pilota progettati per un attacco furtivo in uno spazio aereo contestato.

I classici APR come il Predator ed il Reaper per via delle loro lacune nell'autodifesa e dipendenza dal controllo remoto via satellite (Satcom) presentano molti limiti in caso di un conflitto ad alta intensità[41].

Al contrario, gli UCAV basandosi su istruzioni pre-programmate non presentano questi difetti, anzi, permetterebbero così di assicurare ad uno stato un maggior vantaggio offensivo. Operazioni come: SEAD, OCA e DCA etc. non richiedono, infatti, un alto grado di autonomia in quanto gli UCAV dispongono di munizioni con una significativa autonomia nella distinzione del bersaglio. Proprio questa necessità ha fatto emergere nell'ultimo decennio diversi progetti dimostrativi, che combinano velocità, controlli e comandi sempre più avanzati. Due esempi sono il Taranis e il nEUROn, che possono fondersi come parte del programma di dimostrazione UCAS (*Unmanned Combat Air System*) sostenuto dal governo franco-britannico. Nel marzo 2016 i due governi si sono impegnati a investire congiuntamente £ 1,5 miliardi in uno sforzo congiunto BAE-Dassault guidato dall'Aviazione per costruire un eventuale prototipo.

[41] Il segnale satellitare (satcom) può essere facilmente intercettato, interrotto o falsificato per via della distanza in cui si trova ad operare rispetto alle fonti di disturbo nemiche.

MQ-8C Fire Scout: Definito dalla compagnia come *l'elicottero autonomo di prossima generazione della Marina americana*[42], è in grado di eseguire compiti di intelligence, sorveglianza, ricognizione e acquisizione del bersaglio (ISR&T) tramite un puntatore laserin tempo reale.

A contraddistinguerlo non sono né la sua autonomia di ben 12 ore, né la sua velocità massima di 135 nodi, ma la sua capacità di decollare e atterrare autonomamente su qualsiasi nave o zona di sbarco anche non adeguatamente predisposta. Un miglioramento significativo che ha di fatto aumentato anche la sua portata e la sua resistenza (più del doppio) oltre alla capacità di carico utile (più del triplo) rispetto alla versione precedente MQ-8. Secondo la compagnia, inoltre, l'MQ-8C ha terminato tutti i test di sviluppo.

Caratteristiche tecniche	
Paese	Stati Uniti
Compagnia:	NorthropGrumman
Lunghezza:	12.6 m
Altezza:	3.3 m
Larghezza:	2.7 m
Velocità:	135 nodi
Autonomia:	12 ore
Payload (interno):	227 kg
Venduto a:	US Navy

Tab. 3.11. *MQ-8C Fire Scout.* Fonte: NorthropGrumman, 2019

L'X-47B: il 10 luglio del 2013, ha completato il suo primo test senza incidenti a bordo della USS George H.W. Bush. Progettato per lanci *stealth* e *carrier-based*, il programma X-47 della DARPA, impiega tecnologie già note come il GPS, l'autopilota, sistema di data LINK 16 e sistema anti collisione[43]. La sua caratteristica principale risiede nella capacità di autorifornimento in volo, tecnologia attualmente appannaggio solo della NorthropGrumman, la quale permette così di aumentare non solo il suo *range*, ma anche la sua durata e la sua flessibilità operativa, sebbene sia attualmente progettato solo per compiti di *intelligence*, sorveglianza e ricognizione. A maggio del 2015 il programma per i test di collaudo è stato definito completato, portando il progetto in *stand-by*, ad oggi, non si sa se e quando riprenderà in futuro.

[42]www.northropgrumman.com/Capabilities/firescout/Pages/default.aspx
[43] https://www.cesi-italia.org/articoli/158/x47b-il-futuro-della-us-navy-adesso

Caratteristiche tecniche	
Paese:	Stati Uniti
Prodotto da:	NorthropGrumman
In appalto dal:	2007
Lunghezza:	12 metri circa
Lunghezza alare:	19 metri circa
Range:	oltre 3.890 km
Due scompartimenti interni per le armi:	peso max. 4,500lb (2.041 kg)
Velocità massima:	High Subsonic
Venduto a:	in sviluppo

Tab. 3.12. *X-47B.* Fonte: NorthropGrumman, 2019

nEUROn: Il sistema di combattimento *Stealth* senza equipaggio (UCAV) sviluppato nell'ambito di un programma dimostrativo da 405 milioni di euro è progettato dalla francese Dassault con il supporto anche di Leonardo[44]. Ha visto il suo primo volo il 1° dicembre del 2012, ad Istres in Francia. Con capacità autonoma di attacco aria-suolo, è progettato per portare due bombe da 250 kg a guida laser contenute nella stiva interna. È classificato come il sistema di arma più autonomo attualmente in fase di sviluppo.

Caratteristiche tecniche	
Paese:	Francia
Prodotto da:	Dassault (primo produttore), Alenia, SAAB, EADS-CASA, Hellenic Aerospace Industry (HAI) & RUAG.
In appalto dal:	2007
Lunghezza:	10 metri
Lunghezza alare:	12.5 metri
Peso:	5.000kg – max 7000kg
Velocità massima:	mach 0.8
Autonomia	3 ore
Venduto a:	in sviluppo

Tab.3.13 - *nEUROn.* Fonte: Dassault Aviation, 2019

Taranis: con uno stanziamento di 185 milioni di dollari (PAX, 2017 p. 11), il Regno Unito, nel 2010, ha presentato un prototipo di aereo da combattimento autonomo senza pilota (UCAV), invisibile ai radar. Definito come il *velivolo dimostrativo tecnologicamente più avanzato mai costruito nel Regno Unito*[45]. Il Taranis sarebbe in

[44] L'Italia risulta il secondo maggior partecipante al progetto, con una quota pari al 22%. Altri partecipanti al programma sono: la Francia, che guida il programma attraverso Dassault Aviation, la Svezia, la Spagna, la Grecia e la Svizzera.

[45] Farmer B., *'Taranis stealth drone may see final test flights later this year'*, 2015, Settembre, The Telegraph. Consultabile in: https://www.telegraph.co.uk/news/uknews/defence/11859967/Taranis-

grado di raggiungere un'area predefinita tramite una traiettoria programmata per identificare e mirare automaticamente al bersaglio. Invia i dati alla sua base di partenza, dove le informazioni sono verificate da un operatore, che dà poi il via libero o l'assenso per l'attacco. Può essere controllato via satellite da qualunque parte nel mondo ed è progettato per missioni intercontinentali; se armato, inoltre, è in grado di colpire con precisione bersagli a lunga distanza sia in aria sia sul suolo. La sua progettazione prevede, per di più, due scomparti per il trasporto di bombe e missili, anche se il loro funzionamento, pur nella loro parziale autonomia, per ammissione dello stesso ministro della difesa britannico, è previsto «*solo sotto il controllo di personale militare altamente specializzato*»[46]. Nel 2016, tuttavia il *Times* ha riportato che la società stava procedendo ad incrementare la sua autonomia poiché «*in futuro potrebbe essere necessaria una capacità di attacco autonoma*»[47].

Caratteristiche tecniche	
Paese:	Gran Bretagna
Produttore:	BAE Systems
Lunghezza	12,5 metri
Apertura alare:	10 metri
Venduto a:	in sviluppo

Tab. 3.14. *Taranis.* Fonte: BAE Systems, 2019

Esempio di come anche la Cina e la Russia prestino molta attenzione agli sviluppi in questo campo sono lo ***Sharp Sword GJ-11****: presentato il 1° Ottobre durante la parata militare che ha celebrato i 70° anniversario della stato socialista.* Secondo alcuni osservatori il GJ-11 altro non sarebbe che la versione finale dello stealth Lijian o Sharp Sword, che ha fatto il suo primo volo di prova nel 2013[48], e l'**S-70 Okhotnik**, russo, un simile *drone stealth* d'attacco attualmente in sviluppo e prodotto dalla compagnia russa Sukhoi (PAX 2019b, p. 20).

3.6 Munizioni guidate di precisione

Conosciute anche come "bombe intelligenti", sono proiettili esplosivi che possono correggere in tempo reale il bersaglio iniziale o eventuali errori successivi una volta

stealth-drone-may-see-final-test-flights-later-this-year.html

[46] BAE Systems, *Taranis – looking to the Future*, consultabile in: https://www.baesystems.com/en/product/taranis

[47] Cole C., '*BAE Systems pushing ahead with autonomous drone targeting*', 2016, 11 June, Drone Wars UK. Consultabile in https://dronewars.net/2016/06/11/bae-systems-pushing-ahead-with-autonomous-drone-targeting/

[48]https://www.globalsecurity.org/military/world/china/lijian.htm

lanciati. I modelli più antichi furono sviluppati durante il secondo conflitto mondiale, ma il loro primo utilizzo fu solo nella guerra del Vietnam. Composte da quattro componenti di sistema (*targetting*/guida, sistema di volo, motore e testata), possono essere usate per molteplici scopi: missili balistici, missili *cruiser*, missili anti-nave, anticarro, ecc. (PAX, 2017, p. 11).

AGM-158C LRASM; È un missile *cruise* anti-nave a lunga gittata, armato di un penetratore e di una testata con frammentazione esplosiva, in grado di trovare autonomamente il proprio bersaglio utilizzando il radar per localizzare ed eliminare eventuali minacce in tutte le condizioni atmosferiche[49]. Il LRASM è anche in grado di colpire obiettivi terrestri. «*È progettato per volare per centinaia di chilometri, in grado da solo di sfuggire ai radar e fuori dal contatto radio con il personale di controllo. Il Pentagono sostiene tuttavia che il nuovo missile anti-nave è solo semiautonomo*[50] *e che spetta agli operatori umani la decisione finale. I funzionari dell'Agenzia per i progetti di ricerca avanzata (DARPA) della Difesa del Pentagono, che inizialmente sviluppò il missile, e la Lockheed Martin rifiutarono di commentare su come l'arma decidesse sugli obiettivi, riferendo che l'informazione era classificata*»[51].

Caratteristiche tecniche	
Paese:	Stati Uniti
Produttore:	Lockheed Martin & DARPA
Scopo:	Missile anti-nave, missile cruise
Lunghezza	4,267 millimetri
Testata:	454 kilogrammi
Peso:	953 kilogrammi
Velocità:	High-subsonic
Venduto a:	US Navy, US Airforce

Tab. 3.15. *AGM-158C LRASM.* Fonte: GlobalSecurity, 2019

MIM 104 Patriot: è uno dei primi sistemi ad introdurre l'autonomia, operativo nell'esercito americano già dal 1984. Il sistema di difesa missilistico conta ben quattro funzioni operative: sorveglianza radar, *command and control*, comunicazione e guida del missile[52], che lavorano insieme per rilevare, identificare e sconfiggere missili balistici tattici, missili *cruise*, droni e altre minacce. «*Un ingaggio del bersaglio può essere effettuato in modalità manuale, semi-automatica o automatica. Quando si decide di colpire il bersaglio, la stazione di controllo seleziona la/e stazione/i di lancio e*

[49] https://www.lockheedmartin.com/en-us/products/long-range-anti-ship-missile.html.
[50] In quanto non supporta la selezione del bersaglio, ma solo il suo attacco.
[51] Markoff, J., *Fearing bombs that can pick whom to kill,* 2014, 11 Novembre, The New York Times. Consultabile in: https://www. nytimes.com/2014/11/12/science/weapons-directed-by-robots-not-humans-raise-ethical-questions.html.
[52] Vitali E.R., *Raytheon aggiorna i missili Patriot degli alleati*, 2015, 3 Aprile, Analisi Difesa. Consultabile in: https://www.analisidifesa.it/2015/04/raytheon-aggiorna-i-missili-patriot-degli-alleati/.

i dati vengono inviati al missile. Dopo il lancio, il missile Patriot viene acquisito dal radar» (PAX, 2017, p. 12).

Caratteristiche tecniche	
Paese:	Stati Uniti
Produttore:	Raytheon
Scopo:	Missile Terra aria con scopo contraereo
Gittata:	70 km
Esplosivo:	90 kilogrammi
Velocità:	High-subsonic
Venduto a:	USA, ed altri 20 Paesi

Tab. 3.16. *MIM Patriot.* Fonte: ArmyTechnology, 2019

SeaRAM: è un sistema missilistico di difesa navale, progettato per contrastare diverse minacce. «*Combina, infatti, precisione, un esteso range ed un'elevata manovrabilità del missile RAM per una rapida e affidabile risposta del sistema PhalanxBlock 1B*»[53]. Il SeaRAM è in grado di rilevare, localizzare e coinvolgere in modo autonomo i bersagli nemici, con una maggiore capacità di colpire più bersagli contemporaneamente, attraverso l'uso di una «*radio frequenza passiva automatizzata dual-mode e guida a infrarossi*» (PAX, 2017, p.12). Funziona in modo indipendente dal sistema di combattimento della nave.

Caratteristiche tecniche	
Paese:	Stati Uniti
Produttore:	Raytheon
Scopo:	sistema missilistico di difesa contro minacce subsoniche, supersoniche, missili cruise, droni ed elicotteri.
Velocità:	High-subsonic
Venduto a:	USA

Tab. 3.17. *SeaRAM.* Fonte: Raytheon, 2019

S-400 Triumph: sistema missilistico russo progettato e sviluppato da *Almaz-Antey con una* capacità contro minacce aeree/missilistiche quali ad esempio: UAV, missili balistici e missili cruise[54]. Con una risposta inferiore ai 10 secondi, il sistema può individuare 36 obiettivi contemporaneamente in un raggio che va dai 40 ai 400 km.

[53] https://www.raytheon.com/capabilities/products/searam
[54] https://www.army-technology.com/projects/s-400-triumph-air-defence-missile-system/

Caratteristiche Tecniche	
Paese	Russia
Prodotto da	*Almaz-Antey*
Armato con:	4 missili di diversa gittata: a lungo raggio da 250 a 400 km, a medio raggio da 120 km e a corto raggio da 40 km.
In servizio:	dal 2007 nelle forze armate russe
Venduto a:	Russia, Cina, India, Turchia[55], Arabia Saudita etc.

Tab. 3.18. *S-400 Triumph*. *Fonte:* Air Power Australia, 2019

Ogni batteria, infatti, montata su un vettore, dispone di quattro differenti missili guidati che con la loro diversa gittata possono essere "mixati" per ottimizzare la sua capacità nei confronti della minaccia in arrivo. L'S-400 è in grado di rilevare e intercettare molteplici velivoli: bombardieri strategici, velivoli radar, aeromobili per la guerra elettronica, aerei da trasporto tattici, missili balistici e missili cruise di tutte le dimensioni.

JSM/NSM (Joint Strike Missile/Naval Strike Missile): Il Naval Strike Missile sviluppato da Raytheon in collaborazione con Kongsberg è un missile cruise di quinta generazione, pensato per essere integrato in diversi veicoli da combattimento compreso l'F-35. Ha un riconoscimento del bersaglio autonomo, reso possibile da un *seeker* di immagini a infrarossi. Tra le sue caratteristiche vi è un sistema avanzato di pianificazione del coinvolgimento ed un collegamento di dati di reti bidirezionale con funzionalità di aggiornamento del target, re-target ed interruzione missione. A differenza della maggior parte degli attuali missili guidati, i JSM/NSMnon sono assegnati ad un bersaglio specifico, bensì ad un'area dove avranno il compito di scovare obiettivi predefiniti (PAX, 2019b, p. 17).

Caratteristiche Tecniche	
Paese	America
Prodotto da:	Raytheon e Kognsberg
lunghezza:	3,7 m
Peso:	307 kg
Gittata:	280 km

Tab. 3.19. *JSM/NSM.* Fonte: Defense Industry Daily, 2019

Dual-Mode Brimstone: la maggior parte delle munizioni guidate in realtà fa affidamento su dati pre-assegnati, per trovare, rintracciare e così colpire i bersagli.

[55] Il 12 luglio 2019, Ankara ha ottenuto dalla Russia le prime componenti del sistema, transazione avvenuta a seguito del rifiuto da parte di Obama di vendere i missili Patriot. (Limes, 11/2019). Santoro D., (2019). "*La Turchia guarda a oriente preparando lo scontro con gli USA"* in *Limes*, Novembre, Limes, p 151.

Ovvero, l'autonomia non supporta la selezione del target, ma piuttosto l'esecuzione dell'attacco. Il missile prodotto da MBDA[56], invece, viene identificato dalla Sipricome una delle poche munizioni guidate con un certo grado di autonomia nel selezionare il bersaglio (PAX, 2019b, p.19). A differenza dei normali missili guidati, viene loro assegnato una determinata area, nella quale avranno il compito di individuare bersagli che corrispondono ad un modello predefinito.

Caratteristiche tecniche	
Paese	Europa (varie sedi)
Prodotto da:	MBDA
Ruolo:	missile di precisione da attacco terra-aria
Peso :	50kg
Lunghezza:	1.8 metri
Diametro:	180 mm

Tab. 3.20. *Dual Mode Brimston.* Fonte: MBDA, 2019

Un altro modello degno di nota è: il siluro pesante **MK 48 MOD 6 AT**[57] che con le nuove tecnologie e con continui upgrade è stato progettato per un'efficacia ottimale contro tutti gli obiettivi sia in ambienti costieri sia in acque profonde. E' dotato di un sonar avanzato, combinato con un avanzato di segnale digitale la guida e il controllo basati su software consentono operazioni autonome e tattiche di *fire and forget*, ingaggio simultaneo sia di più bersagli multipli sia di attacchi ravvicinati.

3.7 Veicoli terrestri senza equipaggio - UGV

In grado di operare senza la necessità di un controllo umano, gli UGV sono utilizzati in diverse operazioni quali: la ricognizione, il pattugliamento, il carico e lo scarico di armi ma a differenza dei veicoli aerei e marini, proprio questi presentano i maggiori problemi. Mentre gli UAV e gli UMV operano in ambienti privi di ostacoli, i sistemi terrestri, oltre a questa difficoltà, si trovano ad operare in ambienti imprevedibili.

Ciò spiega come mai per anni gli UGC siano stati ignorati dal settore bellico. I progressi degli ultimi 10 anni, tuttavia, hanno permesso di colmare un gap che sembrava impensabile.

Per rilevare gli ostacoli, quindi, gli UGV devono rilevare la geometria e la composizione del terreno (il tipo di terreno) in qualunque tipo di condizione e/o di

[56]Una joint venture composta dai 3 leader europei nel settore aerospaziale e della difesa: Airbus (37,5%), BAE Systems (37,5%) e Leonardo (25%).

[57]https://www.raytheon.com/capabilities/products/mk48

scarsa visibilità[58]. I veicoli terrestri fanno così molto affidamento sul loro GPS, rendendoli di fatto vulnerabili ad eventuali tecnologie di disturbo del segnale. Nonostante questa lacuna unita alla loro necessità di una supervisione umana, secondo il generale degli Stati Uniti, Mark Milley, gli UGV con una sempre più maggiore autonomia sono destinati a cambiare definitivamente il combattimento terrestre.

La Russia e la Cina, grazie ai loro numerosi progetti (come l'Uran-9 russo), si stanno sempre più imponendo come nuovi *leader* del settore[59].

Gladiator: Concepito e sviluppato negli anni '90, come supporto ai Marines durante le operazioni STOM[60], è un veicolo tattico a 6 ruote e dispone di supporti per mitragliatrici M249 e M240G ed armi d'assalto multiuso. Prevede un'opzione anche per l'utilizzo di armi non letali. Sebbene sia stato sviluppato comprendendo una cabina di controllo per l'operatore, è stato definito come il «*primo robot da combattimento multiuso al mondo*» con il potenziale «*di poter essere aggiornato a modalità semiautonoma ed anche completamente autonoma*» (PAX, 2017, p.13).

Caratteristiche tecniche	
Paese:	Stati Uniti
Produttore:	NREC & BAE Systems
Venduto a:	in sviluppo

Tab. 3.21. *Gladiator.* Fonte: Global Security, 2019

Robattle LR-3: l'UGV della IAI si basa sull'ELTA'sRobotic Kit e può essere equipaggiato con diversi tipi di armi. Pensato per supportare missioni ad ampio raggio come intelligence, sorveglianza, ISTAR, protezione di convogli etc., secondo l'azienda il sistema può operare anche in modo autonomo a diversi livelli nonostante attualmente funzioni a controllo remoto.

[58]Kelly A., Matthies L., Litwin T., Tharp G., *Obstacle Detection for Unmanned Ground Vehicles: A Progress Report*, Conference paper. Consultabile in: https://link.springer.com/chapter/10.1007/978-1-4471-1021-7_52

[59] http://theavion.com/reality-check-robots-with-guns/

[60] La manovra *Ship to Objective* (STOM) è un concetto tattico che può essere applicato a tutti i tipi di operazioni anfibie, ma in genere implica il superamento allo sbarco. Queste operazioni includono assalti, *raid* ed altre operazioni. https://www.mccdc.marines.mil/Portals/172/Docs/MCCDC/Documents/Concepts/STOM%20May%202011.pdf. https://www.globalsecurity.org/military/systems/ground/gladiator.html.

Caratteristiche tecniche	
Paese:	Israele
Produttore:	IAI - Israel Aerospace Industries.
Peso:	7 tonnellate, capace di trasportare fino a 3 tonnellate.
Comunicazioni:	Utilizza tecnologia 4G
Armi in dotazione:	Il robot trasporta il Pitbull, un Remote Weapon Station (RWS) sviluppato da General Robotics.

Tab. 3.22. *Robattle.* Fonte: IAI, 2016

ThemisAdder: Presentato nel 2015, il sistema cingolato è un veicolo terrestre senza equipaggio, provvisto del sistema di armi remote ST KineticsAdder (Singapore Technology), in grado di trasportare diversi tipi di mitragliatrici. Integrato con un sistema di tracciamento video, questo consente il coinvolgimento di obiettivi fissi e mobili[61]. L'ADDER offre, anche telecamere diurne e notturne, un telemetro laser e un sistema di munizione da 40 mm opzionale. «*Nella modalità a controllo remoto, una stazione di controllo consente all'operatore di ricevere i dati del sensore in tempo reale dal veicolo per poi inviare i comandi attraverso un collegamento tattico di dati. Un sistema a controllo autonomo installato nell'UGV provvede ad una verifica autonoma in tempo reale capace sia di evitare gli ostacoli, sia di dirigere il mezzo lungo il percorso/obiettivo desiderato*»[62].

Ben lungi dall'essere un'arma autonoma, l'UGV necessita di due operatori per funzionare ed il suo raggio d'azione rimane ancora limitato. Tuttavia, molti produttori si sono interessati a questo veicolo considerandolo, invece, una piattaforma di test consolidata ed affidabile.

Caratteristiche tecniche	
Paese:	Estonia
Produttore:	MILREM
Lunghezza & Larghezza	2,1 metri ciascuno
Altezza	0,98 metri
Peso:	850 kilogrammi circa
Carico:	può trasportare circa 750 kilogrammi
Venduto a:	in sviluppo

Tab. 3.23. *Themis Adder.* Fonte: ArmyTechnology, 2019

[61] https://www.army-technology.com/projects/themis-hybrid-unmanned-ground-vehicle/
[62] Ibidem.

Mission Master: l'UGV di Rheinmetall rientra in una categoria più ampia. Oltre alle versioni cargo e di sorveglianza, dispone di una vasta gamma di opzioni nel suo arsenale, tra cui una con due lanciarazzi realizzata da Thales con otto missili da 70 mm ciascuno.

Caratteristiche Tecniche	
Paese:	Germania
Produttore:	Rheinmetall
Velocità Massima:	40 km/h
Velocità massima in acqua:	5 km/h
Carico utile:	300kg
Versioni:	Mission Master UGV – S: per la sorveglianza, ricognizione e operazioni di scouting; Mission Master UGV – P: progettata per le attività di protezione perimetrale, di scorta ed acquisizione e impegno del bersaglio.

Tab. 3.24. *Mission Master*. Fonte: Army Guide, 2019

Il Mission Master può funzionare autonomamente per otto ore e ha una velocità massima di 30 chilometri l'ora. Secondo quanto riferito, la Rheinmetall Canada sta cercando di far evolvere il Mission Master *"aumentando sia la sua autonomia sia la sua capacità di operare in "sciami" di UGV per una stessa missione"*(PAX, 2019b, p.27).

La Russia, invece, dispone di modelli come: il robot stealth kamikaze Nerektha (UGV), un piccolo veicolo da combattimento autonomo già in dotazione alle forze russe e che, una volta lanciato, può raggiungere autonomamente un bersaglio in modalità silenziosa per poi esplodere e distruggere carri armati nemici o addirittura interi edifici; l'URAN 9 (UGCV), prodotto da JSC 766 UPTK, un carro armato leggero senza pilota prodotto in massa, dotato di una mitragliatrice calibro 30 mm e missili guidati anti-carro, che è stato schierato in Siria dove i test in condizioni di battaglia hanno portato a un suo ulteriore aggiornamento[63].

Infine, pur se non proprio pertinente a questa categoria, vi è il super carro armato Armata T-14, orgoglio delle forze armate russe. Attualmente, la torretta può essere gestita tramite un telecomando posto all'interno, ma l'intero carro armato è stato pensato per il controllo da remoto. Sputnik, infatti, ha rivelato che gli ingegneri militari russi stanno lavorando per rendere il T-14 un veicolo blindato completamente autonomo[64]. Nel frattempo, Norico, società cinese, sta sviluppando il suo cingolato leggero Cavalry. Dotato di una stazione d'arma a controllo remoto, armata di una

[63]https://www.forbes.com/sites/noelsharkey/2018/11/28/killer-robots-from-russia-without-love/#783974f7cf01
[64] Ididem.

mitragliatrice e un paio di lanciarazzi. Pur con alti livelli di autonomia, mantiene sempre un operatore al comando. Così la società, sta di fatto aprendo la strada nel settore dei veicoli terrestri senza pilota (UGV) sempre più autonomi al Paese (PAX, 2019b, p. 27).

3.8 Veicoli Marini senza equipaggio - UMV

I veicoli marini senza equipaggio possono operare sia sulla superficie (USV) sia sott'acqua (UUV). Sono tecnologie semi-autonome capaci di raggiungere aree non facilmente accessibili all'uomo e dotate di capacità sia offensive sia difensive. A spingere per un loro diffuso impiego è il fatto di essere armi che, operando in un "ambiente vuoto", assicurano il minor rischio ed il più basso impatto in termini di vite umane (UNIDIR, 2015, pp. 3-4). Si pensi all'operazione *Iraqi Freedom* del 2003 dove sottomarini autonomi sono stati impiegati, per la prima volta, per la bonifica da ordigni bellici in UMM QASR Harbor.

Ad esempio la NATO nel settembre 2019 ha effettuato una grande esercitazione in Portogallo con una gamma di sistemi senza pilota anche in mare. L'Alleanza Atlantica ha affermato che i sistemi senza pilota possono essere "*un punto di svolta per affrontare le molteplici minacce nel settore marittimo*"[65].

La maggior parte di questi sottomarini, privi di equipaggio, richiede, inoltre, un certo grado di autonomia a causa dell'impossibilità di utilizzare metodi standard di navigazione (ad esempio il GPS) (UNIDIR, 2015, pp. 3-4). Infine, possono essere utilizzati per una vasta gamma di applicazioni quali: *Mine Counter Measures* (MCM), *Intelligence Surveillance and Reconnaissance* (ISR), *Anti-Submarine Warfare* (ASW) ed unità d'attacco.

AN2-Anaconda: ideato per le operazioni SOCR[66] senza equipaggio, è in grado di trasportare fino a cinque sistemi d'arma. L'Anaconda interamente in alluminio è una nave autonoma Swift che offre *funzionalità avanzate di sorveglianza e ricognizione, identificazione e intercettazione*. Il concetto di base è quello di sviluppare un *Sensor Bot on Water*, in grado di navigare senza l'ausilio di un waypointpre-programmati utilizzando un GPS e dei sensori autonomi. Dovrebbe diventare una «*imbarcazione completamente autonoma con una IA*» ed «*eseguire così manovre tattiche ed evasive in una determinata area per lunghi periodi di tempo, il tutto senza bisogno di alcun intervento umano*» (PAX, 2017, p. 14).

[65]https://www.nato.int/cps/en/natohq/news_168925.htm?selectedLocale=en
[66] Special Operations Craft Riverine (SOCR)

Caratteristiche tecniche	
Paese:	USA
Produttore:	Swiftships, in collaborazione con the University of Louisiana at Lafayette (USA)
Ruolo:	Operazioni SOCR
Venduto a:	in sviluppo

Tab. 3.25. *AN2 Anaconda.* Fonte: NavalDrones, 2019

Sea Hunter: Lanciata nel 2016, la nave[67] è in grado di operare autonomamente, senza il supporto di equipaggio a bordo[68], per lunghi periodi di tempo[69] e percorrere lunghe distanze rispettando le normali regole di navigazione ed evitando le collisioni[70].

È progettata principalmente per scovare sottomarini nemici ASW (*anti-submarine warfare*) e può svolgere anche operazioni quali la posa e la rimozione di mine oppure con funzioni quali la ricognizione e la sorveglianza. Attualmente viene utilizzato come un banco di prova scientifico e tecnologico. Lo scorso anno, il Sea Hunter è stato il primo USV a navigare autonomamente da San Diego a Pearl Harbor, nelle Hawaii[71].

Caratteristiche Tecniche	
Paese:	Stati Uniti
Produttore:	LEIDOS
Ruolo:	ASW
Lunghezza:	132 piedi
Velocità:	27 nodi
Durata/autonomia:	90 giorni circa
Range:	10 miglia nautiche (stimate)
Venduto a:	in sviluppo

Tab. 3.26. *Sea Hunter.* Fonte: NAVAL Technology, 2019

SeaGull: un prototipo considerato come il primo sistema senza pilota al mondo per le missioni ASW e MCM.

Presentato al Singapore Airshow nel febbraio 2016, il SeaGull può eseguire missioni in acque profonde per quattro giorni di fila ad una distanza visiva fino a 100 km. L'USV, gestito dal porto o dalla nave, può essere controllato da un unico *mission control system* (MCS) ed è disponibile in modalità con o senza pilota. Questa imbarcazione può inoltre essere impiegata anche in missioni di *intelligence*, sorveglianza e ricognizione

[67] Un multiscafo comprendente uno scafo principale e due scafi più piccoli (o galleggianti) attaccati ad esso con travi laterali.

[68] https://www.darpa.mil/news-events/2018-01-30a

[69] Si stima sia in grado di navigare per 2 o 3 mesi in autonomia.

[70] https://www.ilcaffegeopolitico.org/41263/la-crescita-dei-sistemi-senza-pilota-nella-us-navy

[71] https://navaltoday.com/2019/02/01/us-navys-sub-hunter-autonomously-navigates-from-san-diego-to-pearl-harbor-and-back/

(ISR), guerra elettronica (EW), sicurezza marittima e idrografia. Secondo la società l'USV *"offre alle marine un vero moltiplicatore di forze che offre prestazioni migliorate per le operazioni navali, riducendo i rischi per la vita umana e riducendo drasticamente i costi di approvvigionamento e operativi"*[72].

Caratteristiche tecniche	
Paese:	Israele
Produttore:	Elbit Systems
Lunghezza:	12 metri
Venduto a:	In produzione

Tab. 3.27. SeaGull. Fonte: Elbit Systems, 2019

Altri progressi in questo settore sono quelli della Kongsberg, società norvegese, che sviluppa da tempo la sua serie di UUV Ramus per la ricognizione e l'eliminazione di ordigni ed esplosivi. La General Dynamics, invece, ha progettato sottomarini tra i quali il Knifefish e il Black Pearl (PAX, 2019 b, p.29).

Al fine di compensare lo svantaggio navale nei confronti degli Stati Uniti, la Russia, mira a produrre dei veicoli subacquei senza pilota. Un esempio è il Poseidon (nome in codice Uran-6), piattaforma autonoma senza pilota a propulsione nucleare con portata intercontinentale (Saalman, 2019, p. 41).

Anche la Cina, mostra interesse in questo campo: infatti, si contano già in sviluppo diversi modelli di veicoli subacquei senza pilota (UUV) e una nuova serie di grandi sottomarini senza pilota (*Extra-Large Unmanned Underwater Vehicles*, XLUUV). Ed a riprova di tutto questo vi è anche il "Progetto 912" un programma classificato, per cui, secondo esperti, il Paese sta attualmente elaborando robot subacquei militari di nuova generazione in tempo per il centenario del Partito Comunista del 2021 (Saalman, 2019, p. 44-45).

Inoltre, la *China Shipbuilding and Offshore International Company* sta sviluppando il suo USV JARI pensato per operazioni: antisommergibile, anti-superficie e anti-area[73].

3.9 Controllo delle frontiere

Con la tecnologia di controllo delle frontiere, una delle principali nicchie nel mercato militare e paramilitare, la robotica diventa una caratteristica fondamentale delle attrezzature per tale scopo. Simili ai CIWS, ma con calibro inferiori, le sentinelle

[72] Seagull – Multi-Mission USV System, *A Naval Force Multiplier Revolutionizing the Dynamics of Underwater Warfare with Unmanned MCM and ASW Capabilities* in: https://elbitsystems.com/media/Seagull_2016.pdf

[73] https://www.defensenews.com/digital-show-dailies/idex/2019/02/18/ china-is-working-on-killer-robot-ships-of-its-own/

robotiche possono essere più semplicemente descritte come armi antiuomo. Ad oggi, essendo molto rare, possiamo identificarne solo tre diversi modelli.

SENTRY TECH: Sviluppato nel 2007, le autorità israeliane lo hanno utilizzato per proteggere i confini lungo la Striscia di Gaza. È una struttura fortificata modificata (*pillbox*), equipaggiata con la stazione di armi a distanza stabilizzata Samson Mini di Rafael, protetta da scudi pieghevoli. Una serie di questi è collegata in rete e combinata con vari sensori, trasmettendo le informazioni a un singolo operatore che agirà dopo l'identificazione e la verifica da parte del comandante. Sono stati progettati non solo per sorvegliare e controllare eventuali sconfinamenti territoriali, ma anche per proteggere da possibili attacchi con razzi di precisione. Non a caso la loro collocazione logistica è stata da Israele fissata lungo la linea di demarcazione di 60 km con la striscia di Gaza (HRW, 2012, p. 15). «*Le funzionalità di aggancio e di tracciamento automatico del bersaglio estremamente precise del sistema, combinate con un accurato meccanismo di stabilizzazione, consentono prestazioni anche nelle condizioni più avverse*»[74]. «*Una volta che i sensori IDF individuano un potenziale bersaglio, l'operatore può verificare o coinvolgere il bersaglio grazie ai sensori giorno/notte elettro-ottici (EO). Le informazioni acquisite dal rilevatore vengono trasferite nel pacchetto elettro-ottico della stazione d'arma, che si sposta verso il bersaglio, consentendo all'operatore di localizzare e tracciare il bersaglio*»[75]. «*Fonti di Rafael affermano che l'azienda sta ora sviluppando anche un sistema autonomo che non richiederà l'intervento umano*» (PAX 2017, pag 15).

Super Aegis II: per la prima volta in funzione nel 2010, è una torretta automatizzata che può essere montata con una mitragliatrice, un lanciatore automatico di granate da 40 mm o un missile terra-aria portatile[76]. Opera con un raggio di rilevamento anche nell'oscurità totale, utilizzando sensori termici IR, telecamera a colori con ingrandimento 30x, illuminatore laser e laser *range finder*[77]. Dispone di rilevamento automatico, tracciamento, puntamento e accensione manuale o automatica. Può percepire, inoltre, se un bersaglio umano trasporta esplosivi o meno sotto i propri indumenti. Progettato inizialmente per operare in maniera autonoma, per includere funzioni manuali è necessario inserire manualmente un input che consenta alla torretta di scattare. Attualmente l'arma non ha modo di distinguere tra amico o nemico.

[74] http://www.rafael.co.il/Module/ImageDownload.aspx?fl=598&cs=EA636747B4F0AD71DB9D518A28937F8A8F79A455
[75] https://www.wired.com/2007/06/for-years-and-y/
[76] http://www.dodaam.com/eng/sub2/menu2_1_4.php
[77] Ibidem.

Caratteristiche tecniche	
Paese:	Corea del Sud
Produttore:	Dodaam System LTD
Range:	3 km (di giorno), 2,2 km (di notte)
Dimensioni:	786 H x 1,120 W x 700D (mm)
Peso:	120 kg senza armi
Armi in dotazione:	12.7 mm calibro standard, 40 mml GL (opzionale), remote builletrecocking system

Tab. 3.28. *Super Aegis II.* Fonte: DoDAAM System, 2019

SGR-A1 Sentry Guard Robot: La Corea del Sud ha investito in questo modello per il potenziale dispiegamento lungo il confine tra la Corea del Nord e la Corea del Sud (DMZ)[78]. La SGR-A1 ha una mitragliatrice da 5,56 mm ed un lanciagranate da 40 mm e rileva gli intrusi tramite sensori a infrarossi[79]. È dotata di rilevatori di calore e di movimento ed attraverso tali strumenti può percepire persone e mezzi nel raggio di 2 miglia e decidere, dietro un comando umano, ogni azione consequenziale.

Caratteristiche tecniche	
Paese:	Corea del Sud
Produttore:	HanwhaTechwin (prima Samsung Techwin Co.) &Korean University
Ruolo:	Sentry Robot
Altezza:	1,2 metri
Peso:	117 kilogrammi
Range:	Circa 2 miglia
Range sensori diurni:	2,5 miglia
Range sensori notturni:	1,2 miglia
Venduto a:	Corea del Sud

Tab. 3.29. *SGR A1 Sentry Guard Robot.* Fonte: GlobalSecurity, 2019

Il robot fisso utilizza un *software* di riconoscimento dei *pattern* per individuare gli umani e può inoltre riconoscerne eventuali movimenti di resa. La macchina ha disponibile sia una modalità supervisionata sia non supervisionata. «*In modalità senza supervisione, la SGR-A1 identifica e tiene traccia degli intrusi nella zona demilitarizzata, finendo per sparare contro di loro senza ulteriori interventi da parte degli operatori*» (PAX 2017, p. 14).

[78] In realtà, in base all'accordo di Armistizio del 53 tra le due Coree, che proibisce lo spiegamento di armi nella zona, non è mai stato messo in campo nella DMZ. Tuttavia, è stato schierato su base sperimentale in Afghanistan e Iraq.

[79] http://www.globalsecurity.org/military/world/rok/sgr-a1.htm

3.10 SWARM

Lo swarm o sciame rappresenta un uso futuristico dell'autonomia nei sistemi d'arma. Più simile a droni che a LAWS, si basa sulla possibilità di utilizzare un gruppo di singoli sistemi che interagiscono ed operano come un collettivo.

Tuttavia, tale premessa non basta a definirlo, anzi, al fine di comprendere ciò, risulta più intuitivo specificare cosa non è. Non è né un particolare sistema, né un sistema multi robot. Quello che lo caratterizza è una "specifica configurazione": infatti, il controllo da parte dell'operatore avviene non su un singolo apparecchio, ma a livello collettivo.

A distinguerlo poi, due elementi essenziali: la "comunicazione interna" e il "coordinamento". Gli attacchi che utilizzano più sistemi contemporaneamente, ma sono privi di coordinamento, non si qualificano come tali. Ciò nonostante, le singole unità (UAS, UGV etc.) o nodiche lo compongono, pur non essendo avanzati, consentono attraverso il coordinamento e la distribuzione delle attività l'esecuzione di missioni complesse.

Generalmente, lo *swarm* viene identificato come una raccolta di sistemi, tanto *semplici* quanto *complessi*, *omogenei* ed *eterogenei*, spesso privi di equipaggio[80], con quattro principali strutture di comando includibili a loro volta in due grandi categorie: *centralizzati* e *decentralizzati* (Verbruggen, 2019, p.4).

I primi, quelli centralizzati, sono divisibili in *controllo generalizzato*, dove un *leader* comanda tutti i singoli nodi e vi è un *coordinamento gerarchico* dove l'allocazione delle attività avviene attraverso diversi livelli. Tra i loro vantaggi vi è quello di trovare soluzioni in maniera più rapida ai vari problemi e con un comportamento più facile da pianificare mentre, come svantaggio, mostrano una maggiore sensibilità ad una eventuale perdita del loro *Leader*[81].

I secondi, invece, quelli decentralizzati, sono distinguibili in *coordinamento per consenso*, dove i nodi decidono collettivamente come coordinarsi per eseguire le missioni, ed in *coordinamento emergente*: qui ogni singolo nodo risponde alle varie unità che lo circondano. Più facili da espandere "in termini di dimensioni", senza nessun punto di errore, possono operare con una larghezza di banda e sono abili a trovare nuove soluzioni ai vari problemi assegnati ed ottenere così risultati molto complessi. Tuttavia, le decisioni che ciascun *nodo* prende si basano su informazioni localizzate anziché su informazioni aggregate a livello globale.

Ancora in una fase di sviluppo, gli *swarm* di tipo militare sono, tuttavia, testati in ambienti in condizioni altamente ottimizzate. Inoltre, le poche dimostrazioni esterne mostrate riducono al minimo la maggior parte dei problemi operativi che, invece, si presenterebbero in battaglia. Nonostante ciò, risulta innegabile che un loro eventuale

[80] Possono prevedere anche sistemi con equipaggio o comandi a sensori ottici.
[81] Inteso come l'unità principale che gestisce e dirige lo swarm.

utilizzo consentirà in futuro tre notevoli vantaggi tattici, infatti: sono scalabili, possono cambiare il numero delle loro unità in base alla missione da compiere; sono adattabili, utilizzabili in varie missioni; robusti, se un singolo nodo fallisce, altri possono subentrare (Verbruggen, 2019, p.3).

Oltre a quanto elencato, si prevede che gli *swarm* avranno anche un basso costo unitario. Particolarmente promettenti, in particolare quelli *decentralizzati* non necessitano di una comunicazione costante con l'operatore, anzi mostrano una maggiore resistenza alle armi elettromagnetiche. Tuttavia tra i limiti che si riscontrano vi è quella imprevedibilità e semplicità che di fatto rende un effettivo controllo umano su questi sistemi particolarmente difficile e complesso.

Quanto, invece, alle modalità in cui un "controllo" può essere esercitato, si riconoscono quattro distinti metodi:

1- tramite il passaggio degli algoritmi che ne specificano il loro comportamento, gli operatori possono passare da diversi pacchetti di algoritmi pre-programmati a seconda del comportamento desiderato (Chakraborty, 2016, p.8);
2- attraverso la modifica dei parametri del loro algoritmo, funzione che permette di cambiare tanto il loro raggio di azione quanto la distanza massima o minima tra i diversi nodi (Chakraborty, 2016, p.8);
3- un controllo remoto di determinati nodi (leader) tramite input o messaggi intermittenti. Il nodo selezionato può così o inviare i comandi agli altri nodi oppure questi potrebbero come reazione rispondere alle sue azioni. Tuttavia ciò rimane possibile solo se non vi sono ritardi nelle trasmissioni del segnale e con continui aggiornamenti sullo stato dello *swarm* (Chakraborty, 2016, p.9);
4- alterando l'ambiente per influenzarne il suo comportamento. In quest'ultimo caso, tramite l'utilizzo di *feromoni virtuali*[82]viene indicato ai nodi di esplorare una determinata area o meno. Un operatore non invia comandi al gruppo per ottenere un determinato effetto, ma anzi il comportamento deriva dalle modifiche nel modo in cui i nodi interagiscono tra loro o dall'ambiente circostante (Chakraborty, 2016, p.9).

Risulta evidente che, qualunque metodo si consideri, rimane pur sempre una reale difficoltà operativa. Problema accentuato sia da quelle dinamiche poco lineari che limitano la valutazione effettiva delle sue prestazioni, sia dalle interazioni tra i diversi nodi che non possono essere percepite dall'operatore durante la loro formazione. Inoltre, comandi non tempestivi rischiano di produrre non solo effetti opposti a quelli sperati, ma addirittura provocare conseguenze negative.

Per un effettivo controllo, un operatore si potrebbe trovare, infatti, nella situazione

[82]Agenti virtuali, dispositivi virtuali o amorphous computing, ovvero, una serie di interazioni che dipendono dalla posizione e persistono nel tempo o svaniscono lentamente, proprio come nel caso dei feromoni.

di dover optare per il concetto di *negligenza benevolenza,* ovvero, l'idea di attendere piuttosto che dare istruzioni, metodo altamente controintuitivo e che sottolinea come gli *swarm* di tipo militare siano poco adatti a eseguire funzioni particolarmente complesse.

Nonostante tutte le problematiche fin qui esposte, diversi Stati hanno già intrapreso svariati progetti dimostrativi volti a sviluppare e migliorare questa tecnologia in ambito bellico. Tra i programmi attualmente in sviluppo da parte dei vari Stati e per i quali si hanno maggiori informazioni, vi sono:

Il programma OFFSET (*Offensive Swarm Enabled Tactics*) della DARPA, che mira a *"utilizzare sciami di 250 sistemi di velivoli senza pilota (UAS) e/o sistemi di terra senza pilota (UGV) per compiere diverse missioni in ambienti complessi come quelli urbani"*[83]. Programma realizzato in collaborazione con la Carnegie Mellon University, la Cornell University, la Michigan Technological University e altri ancora, nonché con start-up come Corenova Technologies Inc (PAX, 2019b, p. 25).

Sempre della Darpa, il programma Gremlins mira a sviluppare UAS a basso costo riutilizzabili, basati su un'architettura che consente l'impiego di diverse tecnologie come *payload* avanzati ed operazioni autonome.

I Gremlins verranno lanciati da aerei militari mentre si trovano ancora al di fuori dal campo nemico; inoltre, una volta completata la missione, un aereo da trasporto C-130 li recupererà in aria e li trasporterà in una base operativa in cui l'equipaggio di terra li preparerà per poi eseguire un'altra operazione[84]. Alla Dynetics è stata aggiudicata la terza fase del progetto[85].

Anche a livello europeo, sono presenti diversi programmi come il ROBORDER, progetto del H2020 con l'obiettivo di sviluppare UAS eterogenei ed autonomi per il controllo delle frontiere, e l'Euroswarm, uno dei tre progetti nel *Pilot Project* (PP), volto a dimostrare e testare l'utilità degli swarm eterogenei per operazioni di sorveglianza come l'osservazione di campi militari ostili (Verbruggen, 2019, p. 10).

Altri progetti sugli *swarm* al di fuori del quadro europeo, ma a dimostrazione di come la R&S sia in pieno svolgimento sono il *Future Combat Air Systems*, gestito congiuntamente da Francia, Germania e Spagna che vede coinvolte società come Airbus e MBDA, ed il progetto *Tempest*, gestito congiuntamente da Svizzera, Italia e il Regno Unito. Entrambi i progetti mirano a sviluppare aerei da combattimento di sesta generazione (Verbruggen, 2019, p. 12).

Infine, anche la Cina sperimenta tecnologie simili. Nel 2017, infatti, la *China Electronics Technology Group Corporation* CETC, società di proprietà statale, ha eseguito il lancio di 119 droni in grado di eseguire sia il decollo sia la formazione in

[83] https://www.darpa.mil/work-with-us/offensive-swarm-enabled-tactics
[84] https://www.darpa.mil/program/gremlins
[85] https://cbrnecentral.com/darpa-awards-gremlins-phase-iii-contract-to-dynetics/11180/

volo. Secondo il CETC, infatti, "l'intelligenza dello sciame" è considerata il nucleo ed il futuro dei sistemi senza pilota[86].

3.11 Conclusioni

I LAWS (*Lethal Autonomous Weapons Systems*), pur se sono già patrimonio di molti arsenali militari, non sono, tuttavia, ancora militarmente operativi; tuttavia, questa tecnologia in un futuro non troppo lontano renderà inesorabilmente lo scontro sui campi di battaglia sempre più rapido e letale.

Ciò nonostante diverse grandi potenze, come mostrato, hanno già iniziato a finanziare e investire su svariati progetti, coinvolgendo così non solo diversi settori tecnologici, ma anche relativi dipartimenti delle università. Mentre Paesi come gli Stati Uniti e alcuni membri dell'Unione Europea mostrano una maggiore trasparenza in materia di difesa nella R&S, altri, invece, come la Russia e la Cina, tendono ancora oggi a celare la maggior parte dei loro progetti.

Oltre a questo, quello che risulta evidente è che difficilmente nessuno di questi attori rinuncerà a tale nuova corsa agli armamenti, per il timore connesso alla consapevolezza che un eventuale gap militare influenzerà inevitabilmente anche un eventuale competizione politica e strategica.

Modelli come l'UCAV X-47B, l'UGV Themis Adder, l'ASW Sea Hunter ed altri ancora, alcuni in sviluppo altri in una fase più avanzata, evidenziano non solo gli sforzi da parte delle grandi potenze, ma mostrano anche l'elevato livello delle nuove tecnologie belliche fin qui raggiunto.

Tuttavia, tra tutti i sistemi presentati, lo *swarm* rappresenta senza dubbio il vero uso futuristico dell'autonomia nei sistemi d'arma. La possibilità di utilizzare un gruppo di singoli sistemi che interagiscono ed operano come un collettivo rappresenta un evidente vantaggio tattico ed un moltiplicatore offensivo/difensivo considerevole per colui che per primo riuscirà a realizzarlo. È l'ennesima riprova della ricerca di sistemi sempre più "autonomi" che per la prima volta nella storia rischiano addirittura di capovolgere il monopolio sulla guerra da sempre detenuto dall'uomo.

[86] https://www.chinadaily.com.cn/china/2017-06/11/content_29702465.htm

ACRONIMI

AFRL= Air Force Research Laboratory
ASW = Anti-Submarine Warfare
ATR = automatic target recognition
BLOS = beyond line of sight
CCW = Convention on Certain Conventional Weapons
CETC = China Electronics Technology Group Corporation
CIWS = Close in weapon system
C-RAM = Counter-Rocket, Artillery, Mortar
DARPA = Defense Advanced Research Projects Agency
DASA = Defence and Security Accelerator
DCA = Defensive Counter Attack
DECANT = Defence Capability for Autonomous and Novel Technologies
DMZ = demilitarized zone
DOD = Department of Defence
DSTL = Defence Science and Technology Laboratory
EO = electro-optic
FPI = Foundation for Advanced Studies
IDF = IsraelDefense Forces
ISR = Intelligence, Surveillance, and Reconnaissance
KAIST = Korea Advanced Institute of Science and Technology
LAWS = lethal autonomous weapon system
LRASM = Long-Range Anti-Ship Missile
MCM = Mine Counter Measures
MCS = Mission Control System
OCA = Offensive Counter-Attack
ONG = organizzazione non governativa
R&S = Ricerca e Sviluppo
SEAD = Suppression of Enemy Air Defences
SRC = Skolkovo Robotics Center
UAS = unmanned aircraft system
UAV = Unmanned Aerial Vehicle
UCAS = Unmanned Combat Air System
UCAV = Unmanned Combat Aerial Vehicle
UGV = unmanned ground vehicle
UMVs = Unmanned Maritime Vehicles
USV = Unmanned Surface Vehicles
UUV= Unmanned Underground Vehicles

Bibliografia

Amoroso D., Sauer F., Sharkey N., Suchman L., Tamburini G., (2018). *Autonomy in Weapon Systems The Military Application of Artificial Intelligence as a Litmus Test for Germany's New Foreign and Security Policy*, Vol. 49, Heinrich Böll Foundation.

Boulanin V., Verbruggen M. (2017). *Mapping the development of autonomy in weapon systems*. Sipri, Novembre.

Boulanin V., (2019). *The Impact of Artificial Intelligence on strategic stability and nuclear risk*, *Euro-Atlantic Perspectives*. Sipri, Ottobre.

Chakraborty N., Kolling A., Lewis M., Sycara K., Walker P., (2016). "Human interaction with robot swarms: a survey". In *IEEE*,Vol. 46, fasc.1: 1 - 19.

Crosston M., (2019). "Quanto è umana l'intelligenza artificiale? in Il Fattore Umano". In *Limes*, Agosto: 195 - 200.

Gallo B., (2019). "I Killer Robots e l'arte della guerra. Le implicazioni della robotica sui campi di battaglia". In *IRIAD Review*, 10, Archivio Disarmo - IRIAD.

Human Rights Watch (2012). *Losing of Humanity, The case against Killer Robots.* IHRC.

Kelly A., Matthies L., Litwin T., Tharp G. (1996). *Obstacle Detection for Unmanned Ground Vehicles: A Progress Report*, Conference paper.

PAX (2017). *Where to draw the line, Increasing Autonomy in Weapon Systems – Technology and Trends*, Novembre.

PAX, (2019a). *Don't be Evil, A survey of the tech sector's stance on lethal autonomous weapons*, Agosto.

PAX (2019b). *Slippery Slope*, *The arms industry and increasingly autonomous weapons*, Novembre.

PAX, *State of The AI* (2019c). *Artificial Intelligence the military and increasingly autonomous weapons*, Aprile.

Petroni F., (2015)."Compensa e domina: Il Pentagono e la Terza Offset Stategy". In *Limes*, Gennaio: 141 - 147.

Petroni F.,(2013). *"Apocalypse drone".* In *Limes*, Novembre: 135 - 142.

Rossi J.C., (2019) "Intelligenza Artificiale e robotica alla guerra". In *IRIAD Review*, 5, Archivio Disarmo – IRIAD.

Rossi J.C., (2016). "La guerra che verrà". In *Sistema informativo a Schede*,11, Archivio Disarmo - IRIAD.

Saalman L., (2019). *The Impact of Artificial Intelligence on strategic stability and nuclear risk*, *East Asian Perspectives.* Sipri, Vol. II, Ottobre.

Santoro D., (2019). "*La Turchia guarda a oriente preparando lo scontro con gli USA".* In *Limes*, Novembre: 149 - 159.

Sitografia

https://www.ai4business.it
https://www.airforce-technology.com
https://www.analisidifesa.it
www.army-guide.com
https://www.army-technology.com
https://www.armscontrol.org
https://www.avinc.com
https://www.baesystems.com
https://www.cesi-italia.org
https://www.darpa.mil
https://www.dassault-aviation.com
https://www.defence24.com
http://www.defenseindustrydaily.com
https://www.disarmo.org
https://dod.defense.gov
http://www.dodaam.com
https://dronewars.net
https://www.flightglobal.com
https://www.globalsecurity.org
https://www.hrw.org
http://ir.kratosdefense.com
https://www.ilcaffegeopolitico.org
https://www.mccdc.marines.mil
http://www.navaldrones.com
https://www.naval-technology.com
http://www.navweaps.com
https://www.newscientist.com
http://www.northropgrumman.com
https://www.nytimes.com
https://www.paxforpeace.nl
http://www.rafael.co.il
https://www.raytheon.com

https://www.sipri.org
http://www.stopkillerrobots.org
http://theavion.com
http://www.uspid.org
http://www.unog.ch
https://www.wbgroup.pl
https://www.wired.com

Cap. 4
Vulnerabilità delle tecnologie informatiche e AWS

4.1 Introduzione

Le tecnologie informatiche e di comunicazione (ICT) rappresentano un chiaro ed esplicito esempio di tecnologie a doppio uso, militare e civile, come dimostrato dal fatto che gli sviluppi iniziali delle comunicazioni via rete e dei relativi protocolli di trasmissione dati (TCP/IP in particolare) furono condotti dal Dipartimento della Difesa degli Stati Uniti attraverso la sua agenzia di ricerca DARPA (Defense Advanced Research Projects Agency) con il pionieristico progetto ARPANET alla fine degli anni '60. Durante la guerra fredda ci fu un uso esteso di supercomputer delle generazioni più avanzate e delle reti di trasmissione a scopi bellici e questa co-evoluzione, pur nelle mutate condizioni storiche, é proseguita fino ad oggi.

Il relativamente recente sviluppo dei sistemi di arma autonomi e la quasi contemporanea esplosione delle tecniche di Intelligenza Artificiale (IA) non possono che avere una inevitabile convergenza verso dispositivi capaci di operare con sempre maggiore indipendenza dal controllo umano, passando dal cosiddetto *man-in-the-loop* (controllo umano diretto), attraverso il *man-on-the-loop* (controllo umano indiretto) fino eventualmente a cercare di raggiungere lo stato di *man-off-the-loop* di una completa autonomia del sistema dal controllo umano. Da notare che, se ci limitiamo al solo mondo digitale questa autonomia, è già stata raggiunta da circa due decenni attraverso l'uso dei cosiddetti *worm*, segmenti di codice in grado di moltiplicarsi e propagarsi autonomamente in rete senza necessità di controllo umano nella ricerca del bersaglio e nel successivo attacco (a differenza dei computer *virus*, non autonomi nella propagazione); i *worm* sono già da tempo utilizzati per scopi di guerra cibernetica.

Ritornando alle tecniche di IA, recentemente e nel giro di brevissimo tempo alcuni algoritmi sono stati in grado di apprendere autonomamente e senza addestramento umano giochi di strategia molto complessi; apparentemente queste tecniche di auto-apprendimento sono giunte ad un grado di generalità molto elevato. Si potrebbe avere superato una pietra miliare nelle tecnologie di IA che sembrano essere in grado di gestire e risolvere autonomamente problemi ben definiti, in un ambiente sufficientemente delimitato, con uno scopo ben preciso ed interazioni simulabili, che permettono ai sistemi di apprendere con il passare del tempo (UNIDIR 2018).

Nel caso particolare dei LAWS subentra un ulteriore importante fattore da considerare: i tempi caratteristici di funzionamento e risposta di tali sistemi, in particolare con funzionalità IA, sono molto più brevi di quelli umani, restringendo drasticamente il tempo del ciclo osservazione-decisione-azione (tipico delle operazioni militari) con conseguenti regole di ingaggio inevitabilmente più rapide ed una velocità

delle operazioni molto più elevata che un umano non sarebbe in grado di affrontare, ancor meno nel caso di sciami di sistemi autonomi che operino collettivamente. Sistemi autonomi di questo tipo possono probabilmente essere contrastati solo con una tecnologia equivalente, oppure con altre opzioni a livello cibernetico, il cui controllo umano in tempo reale risulta comunque estremamente limitato.

È essenziale mantenere sempre ben presente che ci si trova ad affrontare un contesto, sia tecnologico sia di conoscenze, in rapidissima evoluzione e forse con pochi precedenti nella storia, dove le previsioni o le proiezioni nel futuro sono molto difficili ed il "futuro" stesso va inteso a relativamente breve termine. Dal punto di vista strettamente scientifico, ad oggi gli algoritmi di IA, e le reti neurali in particolare, mancano di un fondamento teorico solido, proprio mentre le applicazioni tecnologiche decollano a ritmo incontrollato.

In questo scenario è assolutamente fondamentale cercare di analizzare ad ampio spettro, oltre alle considerazioni puramente tecnologiche, gli associati aspetti legali, etici e strategici, un compito estremamente complicato.

4.2 Vulnerabilità intrinseche delle tecnologie informatiche

L'affidabilità dei sistemi cibernetici imbarcati nei LAWS, con le sue numerose componenti inclusa anche la funzionalità di IA, non può prescindere dalla considerazione che questi sistemi sono costituiti, in larga parte, da programmi software in esecuzione in computer tradizionali. Essi sono pertanto suscettibili di tutte le principali vulnerabilità caratteristiche associate alle tecnologie informatiche e di comunicazione (ICT) che saranno qui di seguito brevemente descritte, allo scopo di tratteggiare il panorama globale o almeno la sua parte più tradizionale.

Innanzi tutto si deve tenere presente che, durante l'evoluzione iniziale e lo sviluppo del futuro ecosistema cibernetico globale avviato in USA negli anni '70, le problematiche di sicurezza digitale furono relativamente sottovalutate, se non quasi trascurate, poiché computer e reti della Difesa degli Stati Uniti erano sotto il completo controllo militare e solamente in seguitosi è avuto il graduale passaggio delle infrastrutture di rete verso il mondo civile e commerciale di cui siamo testimoni oggi. Questa disattenzione e trascuratezza, il basso livello di consapevolezza delle potenziali conseguenze ed il comportamento istintivamente collaborativo dei primi utilizzatori ha consentito e, purtroppo, orientato verso la costruzione di strumenti e di protocolli che in certi casi non possedevano alcune delle caratteristiche e degli strumenti necessari alla protezione digitale.

Questa pesante eredità è tutt'ora molto pertinente e causa di molti dei problemi e vulnerabilità dell'attuale mondo digitale estesosi nel frattempo a livello planetario; ciò ha determinato una situazione di ancor più difficile risoluzione che nel passato a causa

dell'ipersviluppo e della proliferazione delle infrastrutture di calcolo a livello globale, infrastrutture ormai difficili da aggiornare anche solo per il numero elevatissimo e la diversificazione degli apparati installati, creando una grave inerzia a qualsiasi tentativo di rilevante modifica di alcuni basilari protocolli di comunicazione.

Alle problematiche intrinseche di sicurezza indotte da certi protocolli "deboli" (a livello di progettazione o implementazione), va aggiunta la attuale incapacità tecnologica di scrivere software immuni da difetti logici o errori (*bug*), la cui entità è stimata nell'ordine di qualche decina ogni 1000 linee di codice, e che sono la causa primaria delle vulnerabilità comunemente sfruttate nelle tecniche di attacco cibernetico.

Tutto ciò crea un sistema globale dove il cosiddetto *malware* può assumere moltissimi aspetti diversi. *Worm* e *virus* sono segmenti di codice in grado di riprodursi e propagarsi autonomamente in rete e dotati di istruzioni di attacco specifico per il bersaglio prescelto (ad esempio l'esfiltrazione di dati o l'intromissione illecita e il controllo remoto di computer); i *Denial of Service* permettono il sovraccarico, fino a giungere al blocco, di segmenti di rete o di particolari servizi. *Rootkit* e *backdoor* sono in grado di nascondere la presenza di software ostile sui nodi compromessi e permetterne l'accesso permanente non autorizzato a determinate funzionalità. I *botnet* sono grandi reti di computer illegalmente penetrati, presi sotto il completo controllo remoto di organizzazioni fantasma ed utilizzati per vari scopi illegali. Di particolare rilevanza sono le cosiddette vulnerabilità *zero-day*, completamente sconosciute anche agli stessi sviluppatori del software in questione e spesso disponibili ad elevate cifre al mercato nero; le *zero-day* sono utilizzabili per gli attacchi finché non sono creati e distribuiti dai produttori del software gli specifici meccanismi di protezione e possono purtroppo permanere nell'ecosistema digitale anche per anni prima di essere scoperte e corrette.

Questo è solo un brevissimo elenco non esaustivo di meccanismi di attacco, ma l'origine delle vulnerabilità può purtroppo avere anche una origine completamente diversa, con un conseguente aumento dei vettori e della superficie di attacco dei sistemi digitali. Non solo il software, ma anche l'hardware ed il firmware (quale il BIOS) possono contenere vulnerabilità, in questo caso situate ad un livello molto più difficile da rivelare e potenzialmente con un profondo impatto; è evidente come una *backdoor* a livello hardware permetta un accesso completo all'apparato in questione, aggirando completamente i sistemi di controllo e autenticazione.

A fine 2017 sono state scoperte due vulnerabilità di questo tipo che hanno interessato una frazione elevatissima dei microprocessori installati sui computer costruiti durante gli ultimi venti anni. Tutti i più importanti e diffusi sistemi operativi erano esposti, tanto da rendere a rischio di attacco anche apparati mobili, smart TV, stampanti e dispositivi di rete, quindi tutto l'ambiente digitale. Le vulnerabilità erano talmente fondamentali e globalmente diffuse da essere considerate catastrofiche dagli

analisti di *security*.

Spectre e Meltdown (Google, 2018), questi i nomi dati alle diverse varianti di queste vulnerabilità, permettevano ad un attaccante di fare accesso a meccanismi interni dei processori e zone di memoria riservate del sistema operativo operante su computer dotati di specifici, ma diffusissimi, processori Intel, IBM e di altre architetture, zone di memoria che i sistemi operativi assicuravano di tenere assolutamente protette, separate e non comunicabili. Attraverso questo "canale" aperto a livello hardware strutturale, le protezioni a livello software superiore risultavano inutili e inefficaci.

La soluzione definitiva a questo problema, come dichiarato successivamente anche dai produttori, era la riprogettazione delle CPU in oggetto.

Purtroppo questo tipo di problematica non è limitata alle unità di elaborazione centrale (CPU), ma si estende ad altri microcircuiti specializzati inclusi nell'architettura di base dei computer, quali ad esempio i processori grafici (GPU, Graphics Processing Unit), adibiti a velocizzare l'elaborazione delle immagini e la gestione video. In virtù della loro alta efficienza in calcoli paralleli, tali unità sono sempre più utilizzate anche per scopi diversi dalla gestione grafica e ovviamente non sono esenti da vulnerabilità della stessa natura di Spectre e Meltdown, come dimostrato dai numerosi aggiornamenti di sicurezza frequentemente distribuiti dai produttori di tali circuiti.

In proiezione, entrando nel contesto specifico dell'attuale rapida espansione delle tecniche di intelligenza artificiale, si sta sviluppando un'altra classe particolare di circuiti integrati chiamati *Tensor Processing Unit* (TPU) adibiti specificatamente per l'uso delle reti neurali, strutture che sono alla base dei meccanismi di apprendimento dell'IA. Naturalmente questa ulteriore struttura hardware, anche se inclusa in sistemi dedicati, difficilmente sarà esente dallo stesso tipo di rischi.

Questi tipi di vulnerabilità hardware mettono in luce un ulteriore livello di potenziale pericolo di manipolazione o di intrusione nei sistemi cibernetici, quello della cosiddetta *supply chain*, cioè la catena di produzione dei micro-dispositivi hardware (CPU, unità logiche, interfacce, ecc.) inseriti all'interno di computer e apparati di rete.

Se gli apparati in uso non possono essere considerati affidabili perché contenenti particolari sottosistemi elettronici vulnerabili o manipolabili, eventualmente prodotti o modificati con scopi ostili in paesi antagonisti, le conseguenze possono essere molto gravi a vari livelli; immaginiamo solamente la potenziale perdita di controllo di complessi sistemi industriali o di infrastrutture critiche a livello nazionale, sia in un contesto civile sia militare, incluso quello dei sistemi d'arma più avanzati a livello cibernetico, e senza dimenticare il possibile impiego da parte di servizi di intelligence stranieri.

In determinati contesti particolarmente delicati è il meccanismo stesso di outsourcing degli apparati o di alcune specifiche componenti che può essere messo in discussione, o comunque il controllo dettagliato della catena di approvvigionamento che assicuri l'integrità dalla produzione alla distribuzione, per prevenire la

proliferazione di tecniche e strumenti ICT occulti ed ostili.

Ci troviamo di fronte ad uno degli argomenti chiave discussi anche a livello internazionale nel contesto delle ICT, a causa delle importanti conseguenze nell'ambito della sicurezza (ONU, 2015).

Va inoltre considerata sotto questa luce anche la prospettiva della futura dinamica della *Internet of Things* (IoT), la proliferazione della interconnessione in rete di apparati di uso comune, che pone problematiche molto complesse dal punto di vista della sicurezza informatica in generale.

Per completezza si aggiunge infine un'ulteriore dimensione delle tecnologie di attacco, questa volta non connessa alle infrastrutture tecniche, ma mirata a quello che a volte risulta l'anello più debole della catena di difesa, o il più facile da attaccare, e cioè l'elemento umano. Il cosiddetto *social engineering* si inserisce a livello comportamentale umano e include la manipolazione fraudolenta di dati o informazioni trasmesse al soggetto bersaglio (utente o amministratore di sistema); lo scopo, spesso raggiunto, è di modificare e guidare inconsciamente le sue azioni (contro il suo stesso interesse) o divulgare informazioni confidenziali che compromettano la sicurezza globale dell'infrastruttura sotto attacco. Si tratta di una manipolazione psicologica a livello individuale o sociale più attribuibile al contesto della *information-security* che della *cyber-security*, ma, poiché risulta uno strumento spesso molto efficace, deve essere sempre tenuto in considerazione.

4.3 Vulnerabilità specifiche delle tecnologie IA/ML

Nel contesto dei AWS, ed in particolare dei sistemi d'arma letali, la prospettiva di un uso sempre più massiccio e diffuso delle tecniche di Intelligenza Artificiale e degli algoritmi di *Machine Learning* appare una tendenza evolutiva inevitabile nel breve-medio termine, poiché questi algoritmi possono favorire ed estendere l'autonomia dal controllo umano. Purtroppo in questo nuovo spazio che si apre, la panoramica delle vulnerabilità software e hardware precedentemente descritta diventa angusta, troppo limitata per descrivere la complessità del fenomeno: l'IA apre una nuova superficie di attacco precedentemente inesistente e che diventerà sempre più importante e rilevante con la diffusione di queste nuove tecnologie.

Le tecniche di ML o *Deep Learning* (DL), basate su reti neurali più o meno profonde, sono in grado di individuare schemi o regolarità di un determinato fenomeno in esame utilizzando un insieme di dati di addestramento da cui estrarre una sorta di modello, per poi eseguire delle operazioni sulla base di questa conoscenza. Possono quindi essere in grado di fare previsioni o prendere decisioni senza un'esplicita e specifica programmazione algoritmica, in certi casi praticamente impossibile allo stato attuale. Questi sistemi sono usati in varie classi di problemi: tra i più importanti si possono

citare la computer vision (identificazione di immagini, supporto alla guida autonoma di veicoli), la classificazione e il riconoscimento di oggetti, la comprensione e l'interpretazione del linguaggio, il supporto alla diagnosi medica ed alcuni meccanismi cibernetici di protezione o rivelazione di attività anomale. Molti sistemi IA sono usati anche semplicemente come supporto al *decision-making* umano, ma in questo contesto specifico diventa particolarmente rilevante e critica la relazione uomo-macchina in termini di affidabilità. Durante la fase di apprendimento (*supervised learning*) la rete neurale "apprende" e memorizza quali siano i risultati desiderati (o la soluzione) di un particolare problema e in una seconda fase è in grado di replicare tale comportamento su dati successivamente sottoposti al sistema risolvendo il problema stesso. In alcuni casi è anche possibile eliminare la fase iniziale di apprendimento (*unsupervised/reinforcement learning*) poiché queste tecniche sono in grado di "autoapprendere" senza supervisione, soprattutto in sistemi ben definiti e descritti con regole precise e formalizzabili. La più recente evoluzione di queste tecniche è dovuta principalmente alla disponibilità di un elevato ed economico potere computazionale, unito ad una grande abbondanza di dati su cui operare l'apprendimento e le successive analisi ed applicazioni.

Per comprendere la potenza di queste tecnologie si prenda in considerazione il caso del progetto AlphaZero: sviluppato con tecniche di IA e utilizzando TPU *ad-hoc*, senza fase di apprendimento supervisionato da umani, ma semplicemente conoscendo le regole del gioco e giocando contro se stesso, è stato in grado di raggiungere un livello super-umano nel gioco del Go, forse il più difficile gioco di strategia da scacchiera. In realtà Alpha Zero (Science, 2018) è stato progettato per imparare anche gli scacchi e Shogi, un gioco della stessa famiglia degli scacchi, quindi le sue capacità sembrano essere astratte ed applicabili a contesti diversi.

Le reti neurali, sui cui si fonda il ML, sono sistemi altamente non lineari e complessi le cui basi scientifiche sono ancora da determinare con precisione, con la conseguenza che molti di questi modelli sono a tutti gli effetti delle "black box", il cui comportamento sfida la comprensione umana ed i cui risultati o decisioni risultano al momento non interpretabili e inesplicabili. Da questa complessità e indeterminatezza nascono le vulnerabilità sfruttabili da eventuali attaccanti, potenzialmente anche con gravi conseguenze.

Si parla di *Adversarial AI* (o *Adversarial Attacks*) quando un aggressore riesce a determinare un particolare comportamento della *black box* (la rete neurale) ignoto ai suoi stessi sviluppatori, per poi sfruttarlo per i propri scopi. Se si riesce a predire il modello di apprendimento di un particolare sistema, può essere possibile in molti casi manipolare i dati per ottenere o forzare il risultato desiderato anziché le decisioni corrette, o in alternativa creare delle vulnerabilità da sfruttare in seguito.

Un serio rischio in questo contesto è quello del *Poisoning Attack* che può essere messo in atto quando si riesce ad influenzare direttamente il processo di

apprendimento delle reti neurali, come nel caso del *supervised learning*. Il *training data set* viene contaminato da informazioni non pertinenti oppure opportunamente costruite allo scopo di indurre il sistema a commettere errori di valutazione o rispondere in un modo pre-programmato ed errato.

In un caso specifico di auto a guida autonoma, a seguito di questo tipo di attacco un segnale stradale di stop è stato classificato dalla rete neurale come un segnale differente, malgrado sembrasse assolutamente normale all'occhio umano, con tutte le potenziali conseguenze del caso. Un altro esempio riguarda i sistemi anti-spam basati su reti neurali: un "avvelenamento" dei dati di training con contenuti di un certo tipo potrebbe istruire il sistema a considerare legittime delle email che andrebbero invece scartate; questo tipo di attacchi sono particolarmente diffusi nei sistemi di *reinforcement-learning*, che ricalibrano i loro parametri sulla base dei dati analizzati.

Da un punto di vista generale, all'aumentare della pervasività dei sistemi di AI e ML consegue la necessità di comprenderne limiti, vulnerabilità e meccanismi di malfunzionamento. A questo ultimo scopo si possono distinguere due grandi classi: i malfunzionamenti intenzionali, causati da una attività ostile che opera manomissioni esplicite a livello di classificazione, addestramento o algoritmo, ed i malfunzionamenti non intenzionali dove il sistema produce un risultato formalmente corretto, ma completamente inapplicabile, insicuro o incoerente con la realtà, a tutti gli effetti sbagliato. Al momento gli analisti hanno caratterizzato circa una decina di metodologie diverse per provocare malfunzionamenti indotti ed un numero più ridotto per quelli non intenzionali (Microsoft, 2019).

Esempi di malfunzionamenti indotti sono la contaminazione ("*poisoning*") dei dati o delle procedure di training (di cui abbiamo già parlato), l'inserzione di eventuali *backdoor* attivate da dati opportunamente manipolati e configurati sottoposti al sistema allo scopo di ottenere la risposta voluta, oppure la dipendenza o l'inoculazione di segmenti di codice vulnerabile (in questo caso un meccanismo di attacco tradizionale, ma applicato nel nuovo contesto IA).

Tra i meccanismi non intenzionali di malfunzionamento, trattasi di fatto di vulnerabilità interne, si possono includere test incompleti del sistema di ML che non comprendono tutte le condizioni realistiche di funzionamento, oppure l'incapacità del sistema di adattarsi alle condizioni ambientali, in particolare con lente variazioni temporali o semplicemente in presenza di dati di ingresso con alta variabilità o perturbati, come rumore, immagini deformate o sporche ecc.

Appare chiaro quindi che i meccanismi di perturbazione o corruzione di questi sistemi non mancano, da cui la necessità di porre la dovuta attenzione nello sviluppo di opportune contromisure che rendano le tecnologie di IA/ML più robuste e sicure allo scopo di evitare conseguenze negative. Va sottolineato il fatto che eventuali malfunzionamenti e manipolazioni di questi algoritmi possono proiettarsi con un effetto a cascata su servizi, applicazioni o prodotti che interagiscano o dipendano da

essi e di cui esiste già un'ampia casistica nel mondo digitale, con la conseguenza di compromettere a valle un ecosistema ben più vasto di quello specifico IA, che include non solo il dominio cibernetico, ma anche quello fisico.

La comunità scientifica, sia civile sia in ambito militare, sta naturalmente affrontando alcuni degli aspetti più problematici recentemente emersi, in particolare quelli che riguardano la attuale caratteristica tipo *black-box* precedentemente citata, molto rilevante soprattutto nel contesto specifico dei LAWS, connessa con l'opacità delle reti neurali e le sue potenziali conseguenze.

La potenza delle tecniche usate nelle procedure decisionali basate su IA appare considerevole, ma l'attuale incapacità tecnica del sistema di fornire agli stessi progettisti una spiegazione su come sia giunto ad una specifica scelta, può risultare un fattore molto limitante per lo sviluppo futuro.

Attualmente stiamo assistendo, da parte di grandi compagnie ICT, all'uso sempre più esteso dell'IA proprio in questo contesto di *decision-making*, con effetti che già influenzano più o meno direttamente i nostri comportamenti, ma affidare decisioni che possono avere conseguenze dirette sulla vita umana (quali diagnosi mediche o sistemi d'arma letali) ad un sistema che non è in grado di spiegare sé stesso presenta evidenti rischi e pericoli e appare difficilmente accettabile.

Questo tema ha sollevato un dibattito internazionale attualmente in corso che sta esaminando un nuovo approccio verso quella che viene definita *eXplainable Artificial Intelligence* (XAI), allo scopo di sviluppare modelli o architetture più trasparenti ed interpretabili e quindi di conseguenza con un maggior grado di affidabilità, pur cercando di mantenere elevati livelli di performance. Si tratta di un tema di ricerca interdisciplinare di estrema importanza nell'ecosistema IA, i cui sviluppi possono potenzialmente coinvolgere ed influenzare uno spettro di applicazioni molto vasto. Al momento ci si trova in una fase iniziale di definizione e di formulazione del problema in termini sufficientemente precisi da permettere un successivo sviluppo, in un panorama dove probabilmente non è ancora stato approfondito a sufficienza il ruolo dell'essere umano nei meccanismi di spiegabilità (IEEE, 2018). Gli approcci potrebbero essere diversi: dalla modifica strutturale delle attuali modalità di *machine learning* e delle reti neurali, ad un diverso ed innovativo metodo di "*explanation by design*".

In ogni caso, la capacità di interpretazione umana sarà indispensabile per comprendere, gestire e poter responsabilmente fare affidamento sulla prossima generazione di tecnologie emergenti in questo campo. Potremmo insomma stare entrando in una nuova era di applicazioni IA.

Sull'argomento specifico XAI, l'agenzia della difesa statunitense DARPA sta finanziando numerosi gruppi di ricerca e si può solo immaginare che valenza possa avere a livello dei sistemi d'arma autonoma.

4.4 Dibattito Internazionale

Nel 2016 nel contesto della Convention on Certain Conventional Weapons(CCW) delle Nazioni Unite fu stabilito un gruppo di esperti governativi (GGE) per approfondire l'argomento delle tecnologie emergenti nell'ambito specifico dei LAWS, con un focus particolare sul diritto internazionale umanitario (IHL), le cui riunioni si sono periodicamente susseguite almeno fino al momento della scrittura di questo testo.

È apparsa subito evidente la difficoltà di comprendere con precisione il concetto di autonomia, un termine che varia con continuità in virtù di un vasto spettro di capacità tecnologiche e del tipo specifico di interazione uomo-macchina.

Considerata la validità e l'applicabilità del diritto umanitario a tutti i sistemi d'arma, la centralità delle scelte e della responsabilità umana risulta fondamentale, ma le tecnologie imbarcate nei sistemi d'arma, includendo anche l'effettivo livello di autonomia e le eventuali tempistiche di comunicazione, mettono questo presupposto a rischio. La natura dei LAWS, con l'eventuale uso di sistemi di IA/ML, pongono dunque nuove sfide alla conformità di tali sistemi al diritto internazionale (ONU, 2019).

I punti critici da prendere in considerazione sono numerosi: la attuale opacità degli algoritmi di ML e la loro effettiva affidabilità, il potenziale di auto-apprendimento con la capacità di evolvere e ridefinire obiettivi operativi in termini di bersagli e regole di ingaggio in modo indipendente dal controllo dell'uomo, responsabile ultimo. Tutto questo è associato ad una possibile imprevedibilità di comportamento. Sono altresì importanti le modalità dei canali di comunicazione con la catena di comando e controllo responsabile durante le operazioni, che però potrebbe non avere più contatto diretto con il sistema dopo l'attivazione. E naturalmente non va trascurata l'importanza di ulteriori considerazioni etiche correlate.

Questo panorama si sovrappone a quello più ampio degli sciami di sistemi autonomi individuali che interagiscono tra loro ed in grado di operare in modo collettivo verso un obiettivo comune; si tratta dei cosiddetti *swarms* (sciami, ad esempio di droni aerei), attualmente di grande interesse militare, che possono porre un serio rischio di violazione del diritto umanitario (EU Non-Proliferation Consortium, 2019).

Questi aspetti tecnologici del dominio cibernetico sono in realtà presi in considerazione anche nell'ambito più generale di altre iniziative ONU che si occupano dello sviluppo delle ICT nel contesto della sicurezza internazionale, dinamica che ha preso l'avvio sin dal 1998 con una serie ininterrotta di risoluzioni dell'Assemblea Generale (*Developments in the field of information and telecommunications in the context of international security*) e rapporti da parte di vari paesi.

Più recentemente, nel dicembre 2018, l'Assemblea Generale dell'ONU ha stabilito due linee di attività parallele, e sperabilmente convergenti, attraverso la creazione di due entità: un gruppo di esperti governativi (*cyber* GGE), con partecipazione selezionata e limitata a 25 paesi e che non comprende l'Italia, ed un *Open Ended*

Working Group (OEWG) che include tutti i paesi membri delle Nazioni Unite.

Il cyber-GGE si focalizza sulle minacce attuali e future provenienti dal dominio ICT, sull'applicazione del diritto internazionale in questo contesto, sulla proposta di norme e principi internazionali a tal riguardo, sulle possibili misure di *confidence building* e di potenziamento delle capacità tecnologiche per i paesi meno sviluppati; il GGE prevede anche di tenere consultazioni con organizzazioni e attori regionali.

L'OEWG prosegue la linea iniziata nel 1998 su argomenti sovrapposti al GGE, ma in un forum nuovo e più ampio (*multistakeholder*) non limitato ai rappresentanti delle varie nazioni, con periodici meeting consultativi che includono il coordinatore del cyber GGE, ma anche rappresentanti di grandi compagnie del settore ICT, del mondo accademico e della società civile; si prevede che i lavori abbiano termine nell' estate del 2020.

La dinamica complessiva cyber GGE-OEWG dovrebbe concludersi nel 2021 e, come si diceva, i LAWS fanno sicuramente parte anche di questo panorama.

A queste iniziative internazionali si aggiungono altre importanti proposte e contributi indipendenti di diversa origine, con grandi corporation ICT coinvolte tra le quali Microsoft, Siemens e Kaspersky. Nel contesto specifico dell'IA è utile citare una lettera aperta indirizzata alla Convenzione CCW dell'ONU sopracitata, resa pubblica da alcune decine di compagnie del mondo della robotica e dell'intelligenza artificiale, che mette in allarme sulle conseguenze delle tecnologie che loro stessi stanno sviluppando:

"Lethal autonomous weapons threaten to become the third revolution in warfare. Once developed, they will permit armed conflict to be fought at a scale greater than ever, and at timescales faster than humans can comprehend. These can be weapons of terror, weapons that despots and terrorists use against innocent populations, and weapons hacked to behave in undesirable ways. We do not have long to act. Once this Pandora's box is opened, it will be hard to close. We therefore implore the High Contracting Parties to find a way to protect us all from these dangers"(Future of Life Institute, 2017).

La stessa organizzazione ha anche suggerito dei principi generali noti come *Asilomar AI principles* (che dovrebbero guidare il futuro sviluppo di queste tecnologie (Future of Life Institute, Asilomar AI Principles, 2017).

Un altro importante contributo a livello internazionale proviene dall'*International Committee for Robot Arms Control* (ICRAC), che raccoglie un gruppo di scienziati in sostegno ad un uso pacifico della robotica, attivo in particolare attraverso consultazioni informali nell'ambito del gruppo di esperti governativi sui LAWS (CCW); ICRAC ha anche prodotto una dettagliata analisi proprio sul concetto di controllo umano significativo, di cui è firmatario uno degli autori di questo rapporto (G.Tamburrini).

Altri importanti contributi dalla società civile provengono dall'International Panel on the Regulation of Autonomous Weapons (iPRAW), un gruppo interdisciplinare di scienziati che si occupa esplicitamente di LAWS partecipando al dibattito in seno alla

convenzione CCW delle Nazioni Unite.

4.5 Conclusioni

Il recente notevole sviluppo dell'IA ha sicuramente molti aspetti positivi ed in certi casi sembra in grado di risolvere taluni problemi con maggiore efficacia di un essere umano, tuttavia è indubbio che queste tecnologie allo stato attuale mostrino nuove vulnerabilità e limitazioni significative. Quello che manca in modo particolare è la capacità di affrontare un "ragionamento" a livello sufficientemente astratto o la capacità di trasferire l'esperienza acquisita verso compiti differenti.

Si tratta insomma di una "intelligenza", se così possiamo davvero chiamarla, relativamente fragile, limitata e ristretta ad uno specifico settore, da cui consegue che resta fondamentale il ruolo e la responsabilità dell'uomo di decidere come, quando e dove utilizzare questi sistemi in modo appropriato, comprendendone le limitazioni, in particolar modo nel contesto militare.

Gli algoritmi IA/ML, il cui sviluppo è recentemente ri-esploso dopo un passato meno appariscente, mancano a vario livello di interpretabilità, predicibilità, verificabilità e affidabilità; naturalmente si tratta di un dominio in evoluzione molto veloce e quindi è certamente possibile che la situazione migliori nel prossimo futuro, ma questa è la situazione attuale. È stato dimostrato che è possibile alimentare questi sistemi con dati scorretti, manipolati o non equilibrati da vari punti di vista, in modo da indurre un comportamento e risultati errati ed in certi casi pericolosi. La proiezione di questa situazione complessiva sui LAWS deve essere analizzata con attenzione, tenendo in considerazione anche la necessità di un comportamento ragionevolmente sicuro dei sistemi e della responsabilità del fattore umano. Non si deve dimenticare che a volte l'uomo è l'anello più debole nella catena di sicurezza cibernetica, e soprattutto nella prospettiva di un uso futuro sempre più diffuso a livello globale, con l'inevitabile coinvolgimento di aspetti etici.

Tutto ciò ricoprirà un ruolo ancora più importante nel contesto militare: la guerra ha luogo in un habitat antagonistico e ostile dove le capacità di adattamento e flessibilità sono caratteristiche primarie e ciò pone una sfida formidabile per il dispiegamento e l'uso di sistemi d'arma letali dotati di funzionalità IA, in situazioni difficilmente prevedibili e potenzialmente fuori dagli schemi di apprendimento, dove gli insuccessi potrebbero condurre a tragiche conseguenze.

A tutto ciò vanno aggiunti ancora due fattori importanti: la sempre più complessa gestione della relazione uomo-macchina e la velocità alla quale i sistemi cibernetici autonomi possono operare, molto più alta della capacità di reazione e intervento umano in tempo reale; va rivolta una particolare attenzione anche al comportamento collettivo di sistemi autonomi intelligenti, quale lo *swarming* precedentemente citato.

Il dibattito internazionale attualmente in corso e brevemente sintetizzato nel paragrafo precedente si sta sviluppando a diversi livelli, ONU, organizzazioni regionali, contatti bilaterali e multilaterali, e deve assolutamente essere attentamente approfondito ed analizzato nei suoi vari aspetti (UNIDIR, 2018).

Appare indispensabile un ampio confronto sul tema delle tecnologie autonome in campo militare da svolgere a livello interdisciplinare, che includa aspetti etici e legali, coinvolgendo la partecipazione di accademici, militari, diplomatici, esperti legali e membri della società civile. Giungere ad una visione globale che includa diverse prospettive sarà fondamentale per lo sviluppo ed un uso corretto e responsabile di queste tecnologie emergenti (UNIDIR, 2018).

Bibliografia

EU Non-Proliferation Consortium (2019). *The Question of Swarms Control: Challenges to Ensuring Human Control Over Military Swarms*(https://www.nonproliferation.eu/the-question-of-swarms-control-challenges-to-ensuring-human-control-over-military-swarms/).

Future of Life Institute (2017).(https://futureoflife.org/ autonomous-weapons-open-letter-2017/).

Future of Life Institute (2017). *Asilomar AI Principles* (https://futureoflife.org/ai-principles/).

Google (2018). *Reading privileged memory with a side-channel* (https://googleprojectzero.blogspot. com/2018/01/ reading-privileged-memory-with-side.html).

IEEE (2018). "Peeking inside the black box: a survey on explainable artificial intelligence". In *IEEE Access*, vol.6, (https://ieeexplore.ieee.org/document/8466590).

"Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model" (https://arxiv.org/abs/1911.08265).

Microsoft (2019). *Failure Modes in Machine Learning* (https://arxiv.org/ftp/arxiv/papers/1911/1911.11034.pdf).

ONU (2015). *Developments in the field of information and telecommunications in the context of international security*RisoluzioneAssembleaGenerale A/70/174 (https://www.un. org/ga/search/view_doc.asp?symbol=A/70/174).

ONU (2019). *Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems – Addendum*

(https://www.unog.ch/80256EDD006B8954/(httpAssets)/931FC63728F1B052C12584AD004A6628/$file/1919338E.pdf).

Science (2018). "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play". In *Science*, 362: 6419, (https://science.sciencemag.org/ content/362/6419/1140).

UNIDIR (2018). *The Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence – Autonomous Weapon Systems and Cyber Operations.*UNIDIR Resources, n.7 (https://unidir.org/sites/default/files/ publication/pdfs/autonomous-weapon-systems-and-cyber-operations-en-690.pdf).

UNIDIR (2018). *The Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence.* UNIDIR Resources, n.8. Geneva: United Nations Institute for Disarmament Research (https://www.unidir.org/files/ publications/pdfs/the-weaponization-of-increasingly-autonomous-technologies-artificial-intelligence-en-700.pdf).

Cap. 5
L'opinione pubblica e le armi autonome

5.1. Opinione pubblica e uso della forza

Erano passati appena quattordici anni dalla vittoria dell'Occidente nella Guerra Fredda, quando si è affacciata sulla scena internazionale quella che il *New York Times* ha definito la nuova 'superpotenza' internazionale, ossia l'opinione pubblica. Ancora in età "moderna", cioè negli anni Cinquanta e Sessanta del Ventesimo secolo, questa entità veniva descritta da politologi, sociologi ed esperti della comunicazione come volatile, disinformata e sostanzialmente indifferente di fronte ai grandi temi politici, specialmente a quelli di carattere internazionale, in quanto remoti rispetto alle conoscenze e agli interessi dell'"uomo della strada" (Isernia 1996). Invece a cavallo tra i due secoli, nella *società post-moderna*, con il declinare delle ideologie, la propensione a cambiare idea viene valutata positivamente. Quella che ancora cinquant'anni fa era connotata negativamente come 'volatilità' oggi tende a essere letta come flessibilità e libertà di giudizio, cioè come capacità del pubblico di adattare il proprio pensiero alle diverse situazioni, senza più subordinarlo ai vincoli culturali, politici e religiosi di partenza e dell'ambiente di riferimento.

Uno dei temi più sentiti a livello individuale e collettivo è, nelle società industriale avanzate, il diritto alla vita e di conseguenza diventano cruciali tutte quelle decisioni che possono metterlo a repentaglio come accade nel caso degli interventi militari.

Il sempre più agguerrito "scrutinio" (Burk, 1998) dell'opinione pubblica, infatti, coinvolge l'insieme degli attori incaricati della sicurezza interna ed esterna e ha per protagonisti tutti coloro che hanno a disposizione un pc, cioè poco meno della totalità dei cittadini elettori. L' "intrusione" di questi ultimi si configura come un dato politico che può essere anche manipolato dal potere ma che non può essere impedito, così condizionando potentemente la scelta dei governi.

Per uno scopo impopolare come l'uso della forza i governi devono garantire alle proprie decisioni la legittimazione, cioè una risorsa che necessita del consenso dell'opinione pubblica.

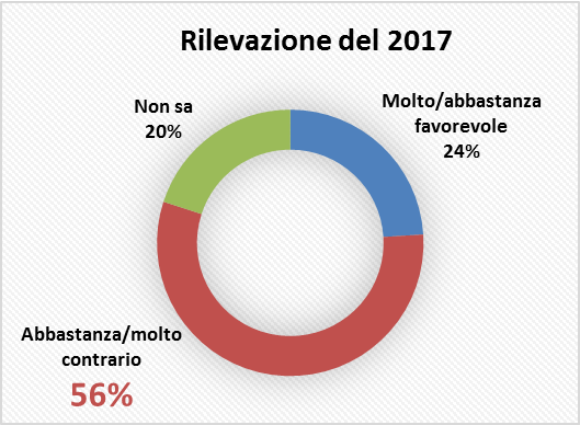
Per ottenere il consenso, o perlomeno l'acquiescenza dell'opinione pubblica è necessario limitare gli effetti letali della guerra, ragione per cui la parola d'ordine vincente è quella delle "perdite zero". A partire dalla "rivoluzione degli affari militari"[87], resa possibile dai progressi tecnologici realizzati nel complesso di capacità ISTAR (*Intelligence Surveillance Target Acquisition and Reconnaissance*), alla fine del Ventesimo secolo, il "sogno" di politici e strateghi sembra avverarsi.

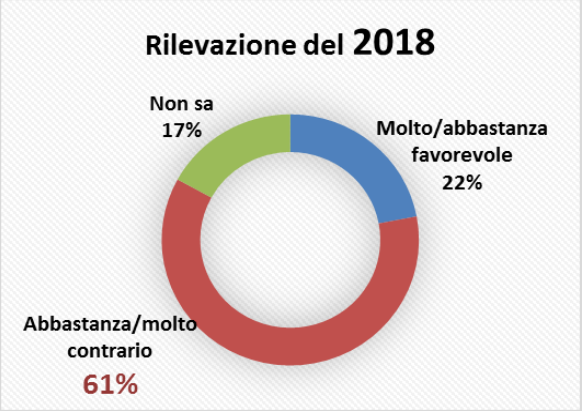
[87] Per l'annuncio datone dal governo americano all'indomani della Guerra del Golfo v. Fiammenghi e Fiorucci, 2009.

5.2 L'opinione pubblica mondiale di fronte alle armi autonome

Nella convinzione che un'eventuale opposizione dell'opinione pubblica allo sviluppo delle armi autonome possa portare i Governi ad agire in modo più deciso verso la costituzione di quadri normativi per contrastare la minaccia proveniente dall'impiego di questi sistemi d'arma, nel gennaio del 2017 la *Campaign to Stop Killer Robots*lancia un primo sondaggio in 23 Paesi volto a rilevare l'atteggiamento dell'opinione pubblica mondiale nei confronti delle LAWS.

L'analisi dei dati, rilevati da IPSOS, ci mostra come il 56% degli intervistati dichiara di opporsi totalmente all'utilizzo dei Robot Killer in guerra, percentuale di oppositori che sale al 61% nella successiva indagine condotta nel dicembre del 2018 coinvolgendo 26 Paesi (v. graf. 5.1)[88].





Graf.5.1. *Favorevoli e contrari alle armi autonome in 23 (2017) e 26 Paesi (2018)*
Fonte: Ipsos per la *Campaign to Stop Killer Robots*
(2017 e 2018)

Da quest'ultima rilevazione emerge come l'opinione pubblica di 22 Paesi su 26 si dichiara contraria all'utilizzo di armi autonome in guerra. I Paesi con un'opinione pubblica maggiormente contraria all'utilizzo delle armi autonome sono la Turchia (78%), la Corea del Sud (74%) e l'Ungheria (74%). Tra le opinioni pubbliche relativamente più favorevoli, invece, si annoverano quelle di India (50%) ed Israele (41%). Quest'ultimo dato non stupisce se si considera come, secondo diverse rilevazioni del *Pew Research Center*, i cittadini di Israele si dichiarano nella misura del 64% favorevoli all'impiego dei droni armati[89].

La più o meno marcata (e comunque maggioritaria) opposizione all'utilizzo delle

[88] La rilevazione dei dati è stata condotta da IPSOS tra il 26 Novembre e il 7 dicembre del 2018 su un Panel Online di 18.795 adulti in 26 Paesi a livello mondiale: Argentina, Australia, Belgio, Brasile, Canada, Cina, Colombia, Francia, Germania, Gran Bretagna, Ungheria, India, Israele, Italia, Giappone, Messico, Paesi bassi, Perù, Polonia, Russia, Sudafrica, Corea del Sud, Svezia e Turchia.

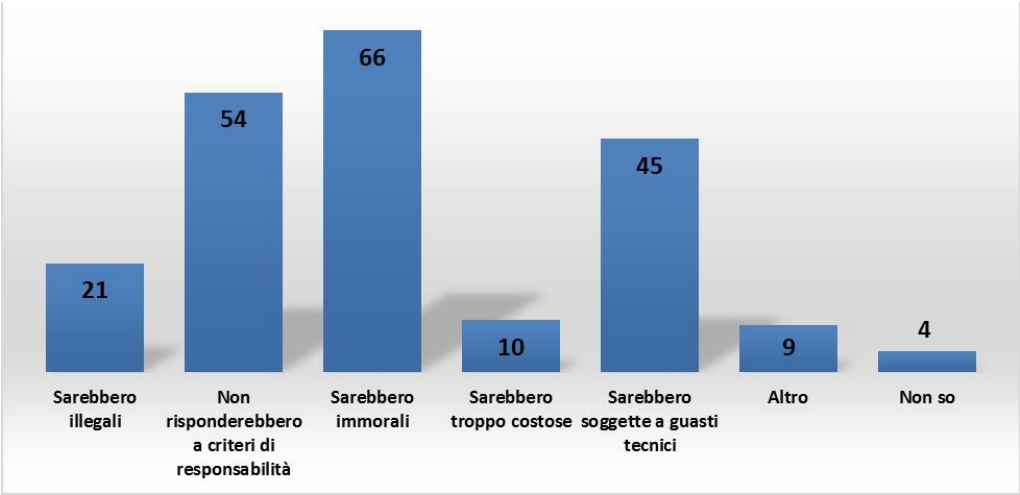
[89]Per un approfondimento su droni e opinione pubblica vedi IRIAD Review 2017 e 2019.

armi autonome in guerra viene dichiarata anche dall'opinione pubblica di Paesi cruciali nello scacchiere mondiale come Cina, Russia, Gran Bretagna e Stati Uniti (v. tab. 5.1). Da notare è come gli statunitensi, favorevoli all'utilizzo dei droni in campo bellico e con un governo che sta ampiamente investendo nella ricerca e sviluppo delle armi autonome, si mostrino contrari all'impiego delle LAWS.

	Molto/abbastanza favorevoli	Molto/abbastanza contrari
Cina	32%	60%
Stati Uniti	24%	52%
Gran Bretagna	21%	54%
Russia	18%	59%

Tab. 5.1. *Favorevoli e contrari alle armi autonome nelle superpotenze mondiali*. Fonte: Ipsos per la *Campaign to Stop Killer Robots, 2018*

A coloro i quali si sono dichiarati contrari alle armi autonome è stato poi chiesto quali fossero le principali fonti di preoccupazione legate al loro utilizzo. I due terzi degli intervistati (66%) ha risposto che i sistemi di arma autonome "supererebbero una linea morale poiché alle macchine non dovrebbe essere permesso di uccidere" e più della metà (54%) sostiene che tali sistemi d'arma non risponderebbero a criteri di responsabilità e di controllo. Rilevante è anche la percentuale di coloro i quali ritengono che le armi autonome sarebbero soggette a guasti tecnici (45%).

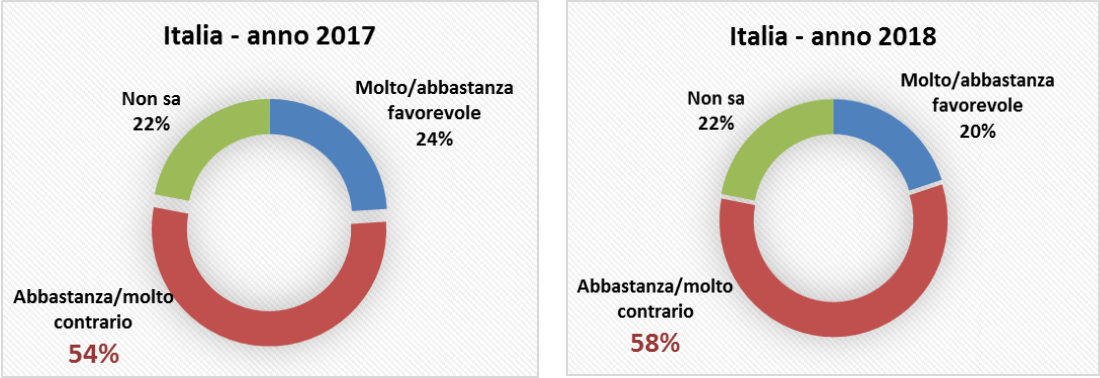


Graf. 5.2. *Principali preoccupazioni relative all'utilizzo delle armi autonome.* Fonte: Ipsos per la *Campaign to Stop Killer Robots, 2018*

5.3. L'opinione pubblica italiana e le armi autonome

5.3.1 L'indagine IPSOS

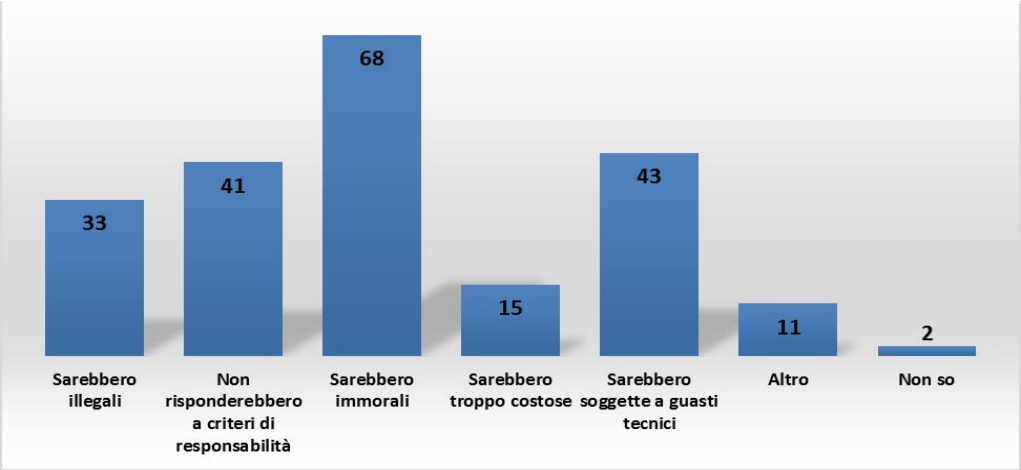
Anche l'opinione pubblica italiana si dichiara in maggioranza contraria all'utilizzo delle armi autonome. Un'opposizione che dal gennaio 2017 al dicembre 2018 cresce dal 54% al 58%.



Graf.5.3. *Favorevoli e contrari alle armi autonome in Italia (anno 2017 e 2018).* Fonte: Ipsos per la *Campaign to Stop Killer Robots*, 2017 e 2018

Come emerso anche dal dato a livello mondiale, la maggior preoccupazione degli italiani rispetto ai sistemi d'arma autonomi è quella relativa al superamento di una barriera morale secondo cui le macchine non dovrebbero poter decidere della vita e della morte di un essere umano. Infatti il 68% degli oppositori a tale sistema d'arma li reputa immorali. L'opinione pubblica italiana, insieme a quella degli altri 25 Paesi coinvolti nella rilevazione, sembra dunque condividere una preoccupazione fortemente espressa da esperti e scienziati a livello internazionale che, oltre ad evidenziare come tali sistemi d'arma potrebbero minacciare le norme legali e diplomatiche stabilite, ritengono che tali macchine non hanno e non avranno mai la capacità morale di decidere della vita o della morte di un essere umano.

Vediamo inoltre come un 43% dei rispondenti è invece preoccupato per i possibili guasti tecnici in cui i robot killer potrebbero incorrere, mentre il 41% teme che il loro utilizzo farebbe venire meno i criteri di responsabilità e controllo umano. Più esigue le percentuali di chi le ritiene illegali (33%) o troppo costose (15%).



Graf. 5.4: *Principali preoccupazioni degli italiani relative all'utilizzo delle LAWS*. Fonte: Ipsos per la *Campaign to Stop Killer Robots*, anno 2018

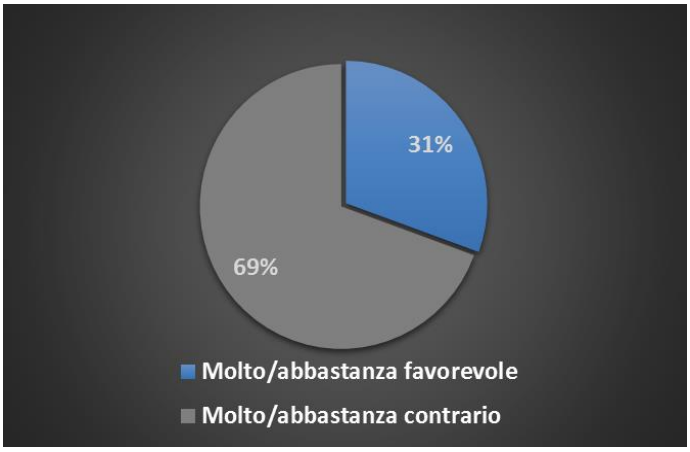
5.3.2 L'indagine di Archivio Disarmo

Nella prospettiva quantitativa Archivio Disarmo ha realizzato un sondaggio di opinione nel quale, seguendo il modello Difebarometro[90], abbiamo sottoposto una domanda sulle armi autonome nell'ambito di un questionario dedicato a un campione di 1000 intervistati di età superiore ai 18 anni, rappresentativo della popolazione italiana per genere, età e residenza geografica.

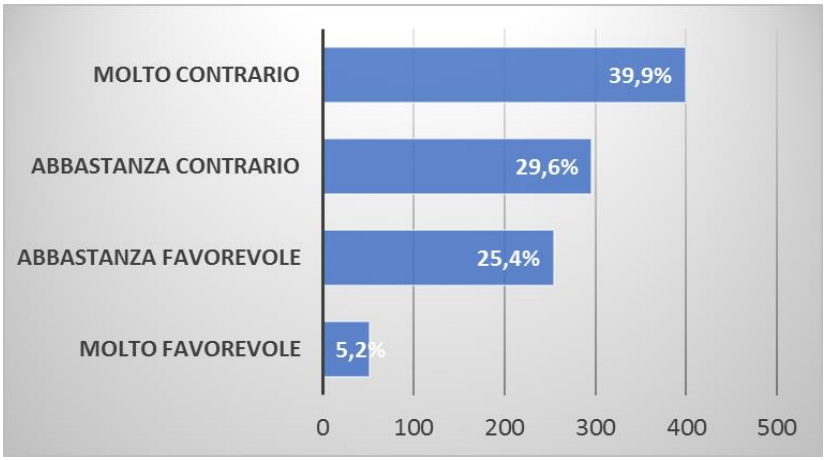
La rilevazione è stata effettuata nel periodo 8-12 febbraio 2019 dalla società Demetra di Venezia, attraverso il metodo mixed-mode CATI - Computer Assisted Telephone Interviewing (50%) e CAWI - Computer Assisted Web Interviewing (50%). In vista dell'analisi dei dati sono state considerate tre variabili: il grado di scolarizzazione, la condizione professionale e la collocazione politica dichiarata dai rispondenti secondo il partito votato alle ultime elezioni.

Richiesti di esprimersi circa lo sviluppo ad opera di alcuni paesi di armi autonome in grado di operare da sole sul campo di battaglia, oltre 2/3 degli intervistati (depurati dei "non sa, non risponde") si dichiarano contrari e un po' meno di 1/3 favorevoli (v. grafico 5.5). Tra i primi, ben il 40% si dichiara "molto contrario" (v. grafico 5.6).

[90]© Archivio Disarmo, febbraio 1995

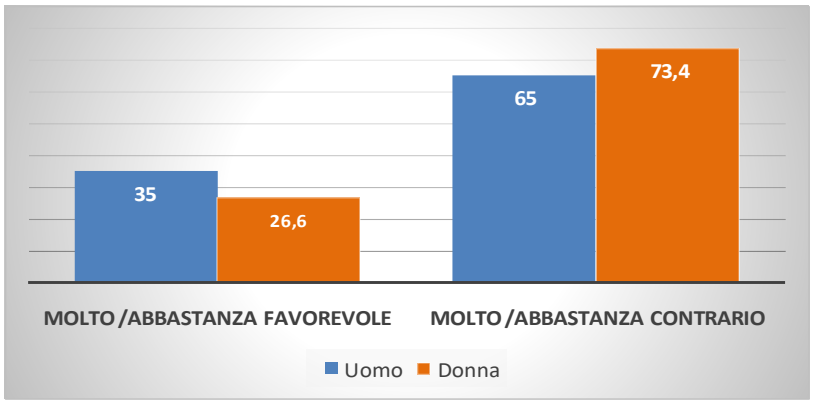


Graf. 5.5. *Favorevoli e contrari alle armi autonome*. Fonte: Rilevazione Archivio Disarmo – Demetra, 02.2019



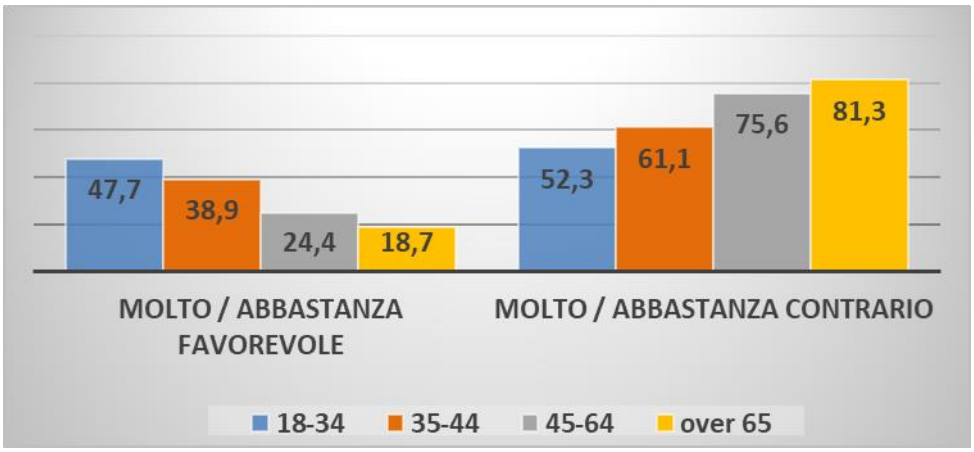
Graf. 5.6. *Gradazioni di favorevoli e contrari alle armi autonome*. Fonte: Rilevazione Archivio Disarmo - Demetra, 02.2019

Come sempre, è interessante correlare l'opinione espressa dagli interessati con alcune caratteristiche strutturali dei medesimi. Iniziando dalla variabile *appartenenza sessuale*, le donne accentuano significativamente la loro contrarietà a questo tipo di armi, nella misura del 73,4% di valutazioni negative rispetto al 65% espresso dagli uomini (v. grafico 5.7). Ne viene confermato il cosiddetto *gender gap*, cioè la tendenza, attestata dalla quasi totalità delle inchieste demoscopiche e dagli studi di opinione pubblica, circa la più spiccata ostilità femminile nei confronti di tutte le forme di violenza, anche quelle istituzionalizzate e legittimamente impiegate dallo Stato per la difesa (Ammendola e Farina, 1999).



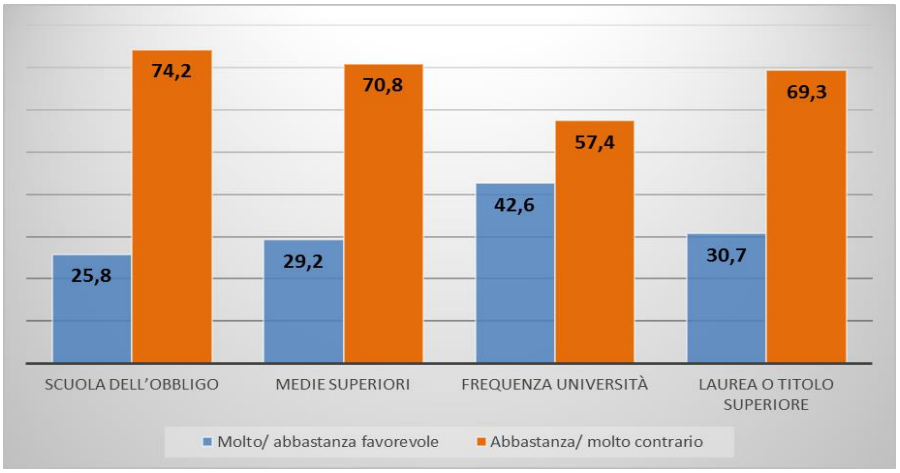
Graf. 5.7. *Armi autonome: favorevoli e contrari alle armi secondo il sesso*. Fonte: Rilevazione Archivio Disarmo - Demetra, 02.2019

Il grado di favore nei alle armi autonome rispetto all'età(articolata in quattro fasce: 18-35, 35-44, 45-64, 65 e oltre) presenta un andamento lineare in cui all'avanzare dell'età corrisponde un atteggiamento via via più contrario nei confronti dello sviluppo delle armi autonome: dalla situazione quasi paritaria tra favorevoli (47,7%) e contrari (52,3%) nella fascia di età 18-34 all'ampia maggioranza di valutazioni negative (oltre 4/5 dei rispondenti) nella fascia di età degli ultra-sessantacinquenni. L'incidenza relativamente elevata di giovani che guardano con favore alle armi autonome appare inversamente proporzionale alla cognizione di ciò che esse potrebbero rappresentare sul campo di battaglia e in genere nella società. Così le evidenti discrepanze tra le varie fasce d'età del nostro campione sono interpretabili con la diffidenza verso l'uso della forza che tende ad aumentare con gli anni degli intervistati e con la presumibile maggiore consapevolezza relativa al ruolo degli armamenti nelle crisi internazionali e nei conflitti. Nel caso dei più giovani esercita un'indubbia influenza la tecnologia come un mito (nel quale possono confluire fascinazioni di diverso spessore, dall'attrattiva per le applicazioni scientifiche alle pratiche consumistiche correnti quali la diffusione dei video game).



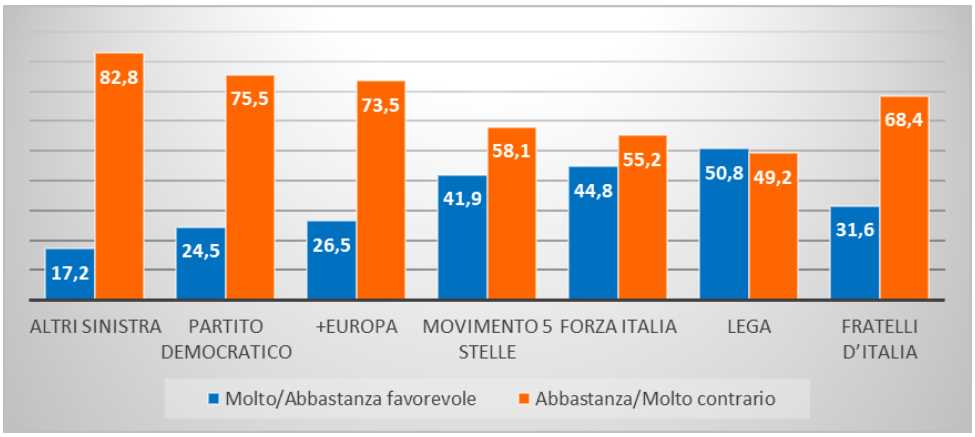
Graf. 5.8. *Armi autonome: favorevoli e contrari secondo l'età*. Fonte: Rilevazione Archivio Disarmo - Demetra, 02.2019

Rispetto ai dati concernenti il sesso e l'età dei rispondenti, il dato relativo alla variabile *istruzione* mostra aspetti in parte controintuitivi (v. grafico 5.9). Mentre il favore verso le armi autonome è minimo presso coloro che possiedono la sola scuola dell'obbligo (26%), esso va aumentando tra chi ha frequentato le superiori (30%) e l'università (43%), tornando infine a calare tra quanti hanno un titolo universitario o superiore (31%). Mentre quest'ultimo dato è in linea con quanto generalmente accertato nelle inchieste demoscopiche (dove a un maggiore livello di istruzione si associa tendenzialmente un maggiore pacifismo), resta da spiegare il picco di coloro che sono (o sono stati) studenti universitari ma non hanno (ancora/mai) conseguito la laurea. Così come ha suggerito la variabile età relativamente alla fascia 18-34, si può ipotizzare che un oggetto del tutto particolare come le armi autonome sia in grado di ispirare sì inquietudine per i suoi risvolti morali, legali e politici ma anche un certo fascino per le sue valenze avveniristiche e fantascientifiche. Quest'ultimo aspetto può influenzare un segmento della popolazione che, per un verso è più smaliziato in quanto studente universitario, ma per un altro non ha maturato l'atteggiamento critico del laureato (ovviamente di età più matura).



Graf. 5.9. *Armi autonome: favorevoli e contrari secondo il livello di istruzione*. Fonte: Rilevazione Archivio Disarmo - Demetra, 02.2019

Infine la variabile *identificazione politica* è esposta nel grafico 5.10 (v.). Rispondendo alla domanda relativa al partito votato alle ultime elezioni, emerge che a condividere sulle armi autonome un giudizio positivo (sia pur di poco: 50,8%) vi sono gli elettori di un unico partito: la Lega. Tutti gli altri sono prevalentemente contrari, con un crescendo di opzioni che ripercorre puntualmente il *continuum* destra sinistra: 55% Forza Italia, 58% Movimento 5 Stelle, 73% Partito democratico 75%, altri sinistra 83%. Un parziale scostamento dalla sequenza in crescita da sinistra a destra è rappresentato da un partito nettamente di destra, Fratelli d'Italia, che con il 68% dei contrari si oppone alle armi autonome in una misura che è a metà strada fra il Movimento 5 Stelle e il Partito Democratico. Fratelli d'Italia è l'erede di una cultura nazionalista che presumibilmente esprime diffidenza nei confronti di strategie e strumenti del modello "americano", presumibilmente preferendo una concezione tradizionale del militare che combatte a viso aperto.



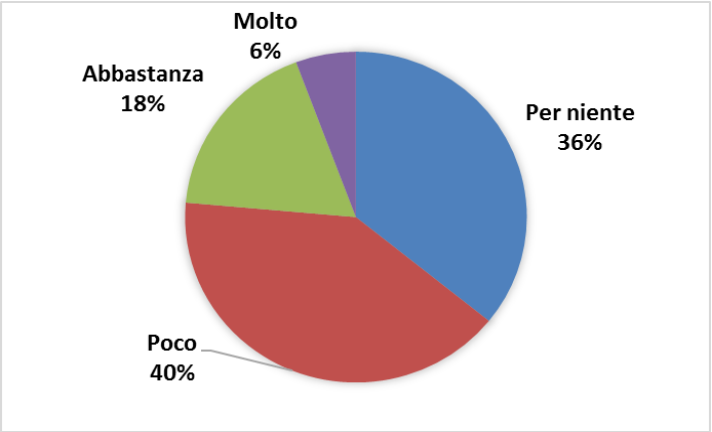
Graf. 5.10. *Armi autonome: favorevoli e contrari secondo il partito votato.* Fonte: Rilevazione Archivio Disarmo - Demetra, 02.2019

In conclusione, almeno in principio l'opzione armi autonome non sembra avere sostenitori (con l'isolata e tenue eccezione degli elettori della Lega salviniana) tra i cittadini italiani. Ciò costituisce una condizione necessaria per l'assunzione di una posizione critica in materia da parte del nostro Paese, che a sua volta potrà divenire condizione sufficiente unicamente se crescerà da semplice espressione di una posizione individuale a presa di coscienza diffusa circa una minaccia che incombe sullo scenario internazionale di qui al prossimo decennio. In tal caso dall'opinione pubblica verrà esercitata una pressione sui media tradizionali e sui social, che a sua volta si trasmetterà su parlamento e governo; questi organi, infine, potranno decidere di portare nelle istanze internazionali (quali i colloqui ONU per la *Convention on Certain Conventional Weapons*) un punto di vista dell'Italia pro-attivo e competente.

5.4 Le giovani generazioni e le armi autonome.

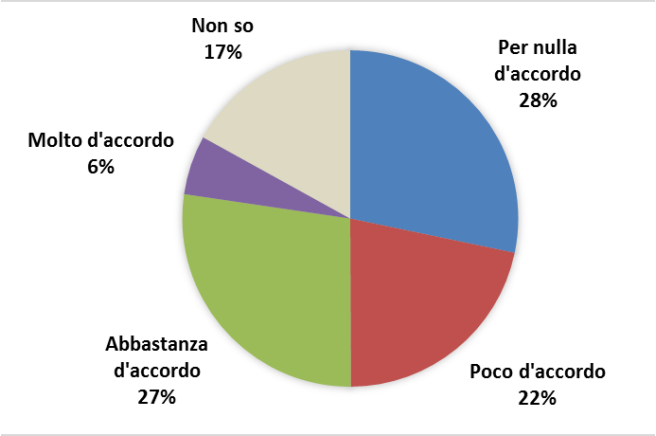
Allo scopo di approfondire che cosa le giovani generazioni conoscono e soprattutto a che cosa pensano quando si parla di armi autonome, abbiamo effettuato uno studio di caso su 106 studenti della Scuola Secondaria di II grado con età compresa tra gli 11 e 14 anni, 62 maschi e 44 femmine. L'indagine ha coinvolto 6 classi di due Istituti comprensivi siti nel Municipio XI di Roma Capitale. La rilevazione, realizzata tra l'ottobre e il dicembre del 2019, si è servita di un questionario strutturato a risposta multipla.

Inizialmente è stato chiesto agli intervistati di dichiarare il proprio grado di conoscenza delle armi autonome. Similmente a quanto emerso da un'analoga indagine avente per oggetto i droni militari (Farruggia, 2018), circa 3/4 dei rispondenti (76%) dichiara di aver sentito parlare "poco o per niente" delle armi autonome; mentre appena 1/4 ne ha sentito parlare o letto (18% "abbastanza", 6% "molto") (v. Graf. 5.11).



Graf. 5.11. *Grado di conoscenza armi autonome*
Fonte: Archivio Disarmo, 2019

Dopo aver brevemente presentato le armi autonome come sistemi di armi che sarebbero in grado di selezionare autonomamente obiettivi e attaccarli senza l'intervento umano, abbiamo chiesto agli studenti intervistati quanto sarebbero stati d'accordo sull'impiego in guerra dei "sistemi letali di armi autonome" in guerra. Non discostandosi molto dal dato a livello nazionale, il 50% si dichiara per nulla o poco d'accordo, contro un 33% che asserisce di essere abbastanza o molto d'accordo. Coerentemente con quanto emerso rispetto al grado di conoscenza delle LAWS da parte degli intervistati, ben il 17% non sa fornire una risposta (v. Graf.5.12).



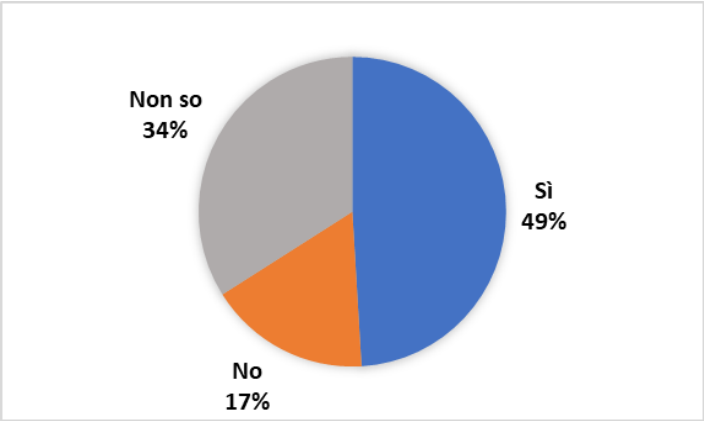
Graf. 5.12. *Grado di accordo su l'utilizzo delle armi autonome in guerra*. Fonte: Archivio Disarmo, 2019

Successivamente è stato chiesto agli intervistati se, a loro parere, l'Italia dovrebbe impegnarsi per un divieto internazionale sui sistemi letali di armi autonome. Poco meno della metà dei rispondenti (49%) ritiene che sì, il nostro Paese dovrebbe impegnarsi maggiormente per una messa al bando delle LAWS, una percentuale formata prevalentemente da coloro i quali si erano dichiarati oppositori rispetto all'utilizzo di tale sistema d'arma; il 17% non ritiene opportuno un particolare impegno da parte dell'Italia; mentre sale al 34% la percentuale di chi non sa esprimere una propria opinione.

Vediamo infine, per coloro i quali si sono dichiarati contrari all'utilizzo in guerra delle armi autonome, quali sono le principali preoccupazioni legate a tale evenienza.

La possibilità di malfunzionamenti tecnici risulta essere il principale timore dei giovani intervistati (35,8%), i quali si dimostrano meno preoccupati dal fatto che l'impiego di armi autonome possa essere illegale (per il 25,5%) o troppo costoso (per il 22,6%). Interessante notare è come la questione morale, risultata essere la prima preoccupazione sia a livello mondiale che nazionale, non assuma grande rilevanza per i giovani intervistati. La possibilità che una macchina possa decidere della vita o della

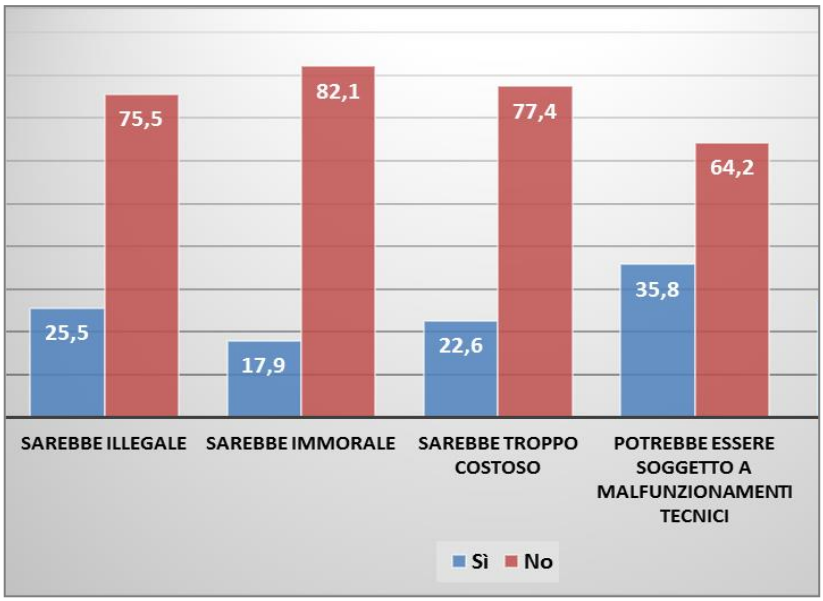
morte di un essere umano sembra impensierire solo il 17,9% dei giovani intervistati.



Graf. 5.13. *Impegno dell'Italia per un divieto sulle armi autonome.* Fonte: Archivio Disarmo, 2019

Vediamo infine, per coloro i quali si sono dichiarati contrari all'utilizzo in guerra delle armi autonome, quali sono le principali preoccupazioni legate a tale evenienza.

La possibilità di malfunzionamenti tecnici risulta essere il principale timore dei giovani intervistati (35,8%), i quali si dimostrano meno preoccupati dal fatto che l'impiego di armi autonome possa essere illegale (per il 25,5%) o troppo costoso (per il 22,6%). Interessante notare è come la questione morale, risultata essere la prima preoccupazione sia a livello mondiale che nazionale, non assuma grande rilevanza per i giovani intervistati. La possibilità che una macchina possa decidere della vita o della morte di un essere umano sembra impensierire solo il 17,9% dei giovani intervistati.



Graf. 5.14. *Principali preoccupazioni relative all'utilizzo delle LAWS*. Fonte: Archivio Disarmo, 2019

5.5 Osservazioni conclusive

L'applicazione dell'intelligenza artificiale all'industria della difesa sta sollevando sempre maggiori preoccupazioni nel dibattito internazionale. In particolare suscitano contrarietà i cd. Robot Killer, sistemi d'arma autonome d'attacco che sarebbero in grado di selezionare e attaccare singoli obiettivi senza un controllo umano. In Italia la previsione di drammatiche conseguenze derivanti dallo sviluppo di tali sistemi d'arma ha portato nel marzo del 2019 più di cento ricercatori di varie discipline (in particolare quelle legate all'informatica, alla robotica e all'intelligenza artificiale) a presentare a governo e Parlamento un appello per la messa al bando delle armi completamente autonome, ritenendole "moralmente inaccettabili".

Le principali analisi sul tema, presentate in questo capitolo, ci mostrano come tali preoccupazioni siano condivise dalla cittadinanza a livello mondiale. L'opinione pubblica di 22 Paesi sui 26 compresi nello studio IPSOS per conto della *Campaign to Stop Killer Robots* si dichiara contraria all'utilizzo di armi autonome in guerra. Gli italiani non fanno eccezione, dichiarandosi "abbastanza" o "molto contrari" nel 58% dei casi (dati IPSOS, 2018), una contrarietà che sale al 68% in una più recente indagine di Archivio Disarmo (febbraio 2019). La principale perplessità, così come emerge anche dal dato a livello internazionale, è la questione morale, in particolare l'inaccettabilità di affidare ad una macchina la decisione sulla vita o sulla morte di un uomo.

Non ci resta che aspettare e vedere se e in che modo a livello mondiale i governi prenderanno atto di questa opposizione pubblica allo sviluppo di sistemi d'arma che sfidano tanto il buon senso quanto il senso comune, sottraendo sia ai decisori (politici) sia agli esecutori (militari) la più delicata e grave delle loro responsabilità.

Bibliografia

Burk J. (1994) (a cura). *La guerra e il militare nel nuovo sistema internazionale*. Tr. it. Angeli: Milano.

Fiammenghi D. e A. Fiorucci (2009). "La rivoluzione degli Affari Militari: perché è avvenuta e che cosa significa". In *Teoria Politica*, n.1, pp. 145-162.

Farruggia F. (2018). "Droni e opinione pubblica. L'analisidei focus group". In *IRIAD Review,* n.8, pp. 21-19.

Isernia P. (1996). *Dove gli angeli non mettono piede. Opinione pubblica e politiche di sicurezza in Italia*. Angeli: Milano.

Cap. 6

Il diritto internazionale umanitario e la sfida delle armi autonome all'*intus-legere*

6.1 Premessa

Dal punto di vista giuridico, sistemi d'arma autonomi non sollevano necessariamente obiezioni: sono molteplici già oggi i sistemi d'arma autonomi utilizzati perlopiù con scopi difensivi. Gli attuali sistemi d'arma autonomi – di contraerea e anti-missilistici – hanno la caratteristica di essere difensivi – reagiscono a seguito di un'intrusione nello spazio aereo con una risposta immediata e limitata – e con un alto grado di prevedibilità nell'esito della loro azione (Iaria, 2018). L'*Iron Dome* sviluppato da Israele, il *Goalkeeper Close-In Weapon System* dei Paesi Bassi o il *Nächst bereichschutz system* (NBS) MANTIS utilizzato dalla Germania sono esempi di sistemi di arma autonomi sviluppati con scopi difensivi e già impiegati che non sollevano alcun dubbio di natura etica o giuridica proprio per la loro caratteristica di essere sistemi d'arma autonomi difensivi e per la loro prevedibilità nel comportamento.

Per molti esperti, il cambiamento chiave che si è avuto negli ultimi anni – una sfida fondamentale nella prevedibilità del sistema – è l'ulteriore sviluppo dell'intelligenza artificiale, e in particolare degli algoritmi che incorporano l'apprendimento automatico. "L'etimologia della parola intelligenza si fa risalire all'avverbio latino *intus* (dentro) ed al verbo latino *legere* (leggere), inteso come comprendere, raccogliere idee e informazioni riguardo a qualcuno o a qualcosa. Quindi, l'intelligenza è la facoltà di comprendere la realtà non in maniera superficiale ma, andando oltre, in profondità, per coglierne gli aspetti nascosti e non immediatamente evidenti" (Etimo Italiano). Nel campo della cibernetica, per il Dizionario Treccani, l'intelligenza artificiale è "la riproduzione parziale dell'attività intellettuale propria dell'uomo – con particolare riguardo ai processi di apprendimento, di riconoscimento, di scelta – realizzata o attraverso l'elaborazione di modelli ideali, o, concretamente, con la messa a punto di macchine che utilizzano per lo più a tale fine elaboratori elettronici". In generale, i sistemi di apprendimento automatico possono avvenire solo in un determinato momento. Il *comportamento* dell'algoritmo di apprendimento è determinato non solo dalla programmazione iniziale (eseguita da un essere umano) ma anche dal processo in cui l'algoritmo stesso *impara* e si sviluppa per *esperienza*. Si individuano dunque due fasi: la prima è data dall'apprendimento tramite la formazione prima dell'utilizzo, mentre la seconda fase è data dall'apprendimento online per esperienza, durante lo svolgimento di un'attività (ICRC, 2018). Armi autonome che integrano l'intelligenza artificiale, permettendo alla macchina di poter imparare (*deep learning*) e quindi di modificare il proprio comportamento, potrebbero non permettere a chi ne sceglie

l'utilizzo di prevederne il comportamento e, in ultima istanza, di essere responsabile della condotta posta in essere.

6.2 Mezzi e metodi di combattimento

Nel corso degli anni la comunità internazionale ha limitato o in alcuni casi vietato alcuni mezzi e metodi di combattimento. Il divieto, per esempio, di ordinare che non vi siano sopravvissuti o il divieto della perfidia, considerate norme di diritto consuetudinario, vietano questi metodi di combattimento. Allo stesso modo, per quanto riguarda i mezzi di combattimento, gli Stati hanno via via regolamentato e messo al bando l'utilizzo di certe armi. In alcuni casi con dei trattati *ad hoc* – Convenzione sulle armi biologiche (1972), Convenzione sulle armi chimiche (1993), Convenzione sulle mine anti-uomo (1997), Convenzione sulle armi a grappolo (2006), Convenzione per la proibizione delle armi nucleari (2017) – in altri casi invece approvando dei protocolli addizionali come nel caso dei cinque protocolli alla Convenzione su certe armi convenzionali del 1980.

L'articolo 35 del Protocollo Addizionale (I) alle Convenzioni di Ginevra stabilisce al primo paragrafo che: «In ogni conflitto armato, il diritto delle Parti in conflitto di scegliere metodi e mezzi di guerra non è illimitato». Limitare la potenza distruttiva e le conseguenze delle armi utilizzate sono concetti che trovavano dei sostenitori già nel 1600. Già Grozio nel 1625 nel suo *De iure belli acpacis*sos teneva la necessità di imporre delle limitazioni sul potere distruttivo delle armi utilizzate in quello che lui definiva come *temperamentum belli*. Un'ulteriore formulazione di tale principio si ha nell'articolo 22 del Regolamento dell'Aja del 1907 che stabilisce come il diritto dei belligeranti di adottare mezzi per ferire il nemico non sia illimitato.

Già SunTzu (544 A.C. – 496 A.C.) nei suoi scritti sanciva come la possibile neutralizzazione del nemico non dovesse per forza coincidere con l'uccisione del nemico (SunTzu, VI Sec. a.C.). Sant'Agostino (354-430) affermava che "se il nemico che combatte deve perire, che questo avvenga per necessità non per tua volontà […] il vinto e il prigioniero hanno diritto alla compassione" (Messineo, 1971), o ancora Francisco de Vitoria (1496-1546), nel suo *De iure belli* scriveva che "non è mai lecito uccidere l'innocente in quanto tale e intenzionalmente, tuttavia incidentalmente, anche se consapevolmente, è lecito in certi casi uccidere innocenti. Si deve evitare che dalla guerra derivino mali superiori a quelli a cui la guerra pone rimedio" (de Vitoria, 1539). Ancora, il giurista Alberico Gentili (1552-1608) affermava che la necessità militare può rendere legale qualunque cosa. Ma l'abuso dei mezzi militari potrebbe indebolire la causa della giustizia e dare all'avversario ragione per muovere a sua volta una guerra giusta. Quindi, non violare le donne del nemico né provocare offese contro la giustizia naturale. E infine Rousseau nel *Contratto Sociale* del 1762, stabiliva come "la guerra è una relazione […] tra Stato e Stato e gli uomini sono nemici solo

accidentalmente […] come soldati […] Lo scopo della guerra essendo la distruzione dello Stato nemico, si ha il diritto di ucciderne i difensori finché sono armati, ma appena questi posano le armi e si arrendono, cessano di essere nemici o strumenti del nemico e ritornano ad essere semplicemente uomini, la cui vita nessuno ha il diritto di prendere" (Rousseau, 1762).

Il già citato articolo 36 del Protocollo Addizionale (I) pone un obbligo nei confronti degli Stati di verificare già durante lo sviluppo di nuovi mezzi e metodi di combattimento se questi siano contrari al diritto internazionale umanitario esistente. Come sottolineato nel commentario del 1987 ad opera del Comitato Internazionale di Croce Rossa, sulla base di questo articolo, le Alte Parti contraenti si impegnano a determinare la natura eventualmente illecita di una nuova arma, sia per quanto riguarda le disposizioni del Protocollo, sia per quanto riguarda qualsiasi altra norma applicabile del diritto internazionale. La determinazione deve essere effettuata sulla base del normale utilizzo dell'arma come previsto al momento della valutazione. Se queste misure non vengono prese, lo Stato sarà in ogni caso responsabile per eventuali danni illeciti che ne conseguono. È sulla base di questo criterio che venne approvato il quarto protocollo addizionale alla CCW, relativo al divieto d'utilizzo dei laser accecanti, prima ancora che questi venissero impiegati. Il divieto di utilizzo di una nuova arma in anticipo rispetto al suo utilizzo venne definito dal Comitato Internazionale della Croce Rossa come «un passo storico per l'umanità. Per la prima volta dal 1868, quando l'uso di proiettili esplosivi fu bandito prima ancora del loro impiego, un'arma di interesse militare è stata bandita prima del suo uso sul campo di battaglia e prima che un flusso di vittime desse prova visibile dei suoi effetti tragici» (ICRC,1995, p. 3).

L'obbligo imposto dall'articolo 36 del Protocollo Addizionale (I) si applica anche per i sistemi d'arma completamente autonomi. L'organizzazione non governativa Article 36 si batte affinché i sistemi d'arma autonomi implementino un «significativo controllo umano negli attacchi individuali»[91]. Questa nozione è stata ripresa da diversi esperti e in effetti sembra rappresentare il requisito sostanziale per far avanzare il dibattito intorno ai LAWS tenendo conto della necessità di prevedibilità dell'arma, e di responsabilità della condotta per garantire la protezione dei civili, discussa precedentemente, e accertare l'eventuale violazione del diritto internazionale umanitario.

Il punto è dunque comprendere quali attività svolte autonomamente dai LAWS debbano essere oggetto della nostra attenzione e dove risulta centrale la prevedibilità della condotta. Un approccio interessante per analizzare il fenomeno è farlo attraverso il c.d. ciclo OODA, osservare, orientare, decidere e agire, ponendosi l'obiettivo di comprendere, in un'operazione militare, quale di queste quattro azioni deleghiamo ai computer. Nelle attività di osservazione e orientamento non si riscontrano profili

[91]La definizione utilizzata in inglese è «Meaningful human control over individual attacks».

direttamente correlati al diritto internazionale umanitario, ma per quel che riguarda la decisione e l'azione risulta cruciale l'elemento della prevedibilità dell'esito tra la decisione e l'azione intrapresa. Se è ragionevolmente possibile prevedere la conformità al diritto internazionale umanitario, si può considerare efficace il controllo umano sull'attività, a prescindere dal livello e tipo di interazione con la macchina al momento dell'azione letale. Se, d'altra parte, non si potrà ragionevolmente prevedere se la macchina rispetterà il diritto internazionale umanitario, l'azione autonoma sarà da ritenersi potenzialmente illecita (Schuller, 2016).

Tale approccio spinge a considerare i sistemi d'arma autonomi, per quanto complessi, ancora come macchine piuttosto che agenti e ancor di più, come agenti legali. Diplomatici ed esperti sollevano dei dubbi sul fatto che i sistemi d'arma autonomi saranno in grado di applicare le norme del diritto internazionale umanitario. Tuttavia se continuiamo a considerare i LAWS come delle macchine, non possiamo che constatare come queste non applichino norme di diritto; possono svolgere funzioni e azioni conformi a quanto previsto dal diritto, ma così facendo non stanno applicando la legge. Il professore Marco Sassòli, nella sua presentazione alla riunione di esperti del CICR del 2014 sulle armi autonome, ha affermato, in modo condivisibile, che «solo gli esseri umani sono destinatari del diritto internazionale umanitario» (Moyes, 2016, pp. 46-52).

6.3 La protezione dei civili

La protezione dei civili rappresenta uno dei due cardini in cui si impernia il diritto internazionale umanitario. Se da una parte infatti questa branca del diritto regola i mezzi e i metodi di combattimento tra le parti in conflitto, dall'altra si adopera affinché le persone che non prendono parte, o non prendono più parte al conflitto, siano tutelate e non siano oggetto di violenza bellica.

Gli Stati assolvono all'obbligo di proteggere i civili attraverso il rispetto, *inter alia*, dei principi di distinzione, proporzionalità e precauzione nell'attacco. Tali principi sono considerati norme di diritto consuetudinario e il loro rispetto è obbligatorio per tutte le parti in conflitto.

Il principio di distinzione obbliga le parti in conflitto a distinguere sempre tra civili e combattenti. Gli attacchi possono essere diretti soltanto contro i combattenti. Gli attacchi non possono essere diretti contro i civili. Sebbene già la Dichiarazione di San Pietroburgo del 1868 stabilisca che l'unico obiettivo legittimo che gli Stati hanno durante un conflitto è quello di indebolire le forze armate nemiche, è il Regolamento dell'Aja del 1907 che, nel proibire l'attacco, con qualsiasi mezzo, a città, villaggi, abitazioni o edifici indifesi, formula una prima base nel riconoscere il principio di distinzione. Dall'obbligo di distinzione tra civili e combattenti, discende il divieto di

attacchi indiscriminati.

La formulazione giuridica del principio di proporzionalità nell'attacco è contenuto nell'articolo 51 paragrafo 5 lettera b del I Protocollo addizionale alle Convenzioni di Ginevra del 1977. Gli Stati sono obbligati a scegliere mezzi e metodi di combattimento che nel perseguire un vantaggio militare concreto e diretto non risultino eccessivi in relazione ai danni collaterali provocati a beni civili o alle persone che non prendono, o non prendono più parte, alle ostilità. Il principio di proporzionalità nell'attacco, dunque, non consiste in un rapporto tra i mezzi utilizzati e i fini raggiunti, quanto piuttosto tra i mezzi utilizzati e il fine ultimo di non causare danni collaterali eccessivi rispetto al vantaggio militare diretto e concreto conseguito. Infine, il principio di precauzione nell'attacco impone agli Stati che nella condotta di operazioni militari si tenga sempre conto della popolazione civile e dei beni civili che potrebbero essere coinvolti durante l'operazione e di prendere tutte le precauzioni possibili per minimizzare il rischio che beni e popolazione civile possano rimanere coinvolti.

La tutela che gli Stati hanno riconosciuto alla popolazione civile attraverso il rispetto di questi principi da parte degli Stati impone che l'eventuale utilizzo di sistemi d'arma autonomi sia conforme ai principi di distinzione, proporzionalità e precauzione nell'attacco. Rimane tuttavia da sciogliere il nodo legato alla responsabilità della condotta.

Secondo il diritto esistente, uno Stato è responsabile per la condotta e le eventuali violazioni del diritto internazionale umanitario commesse da sistemi d'arma autonomi; infatti, l'utilizzo di questo tipo di armi ricade nella scelta e, in ultima istanza, nella responsabilità da parte dello Stato. Inoltre, gli Stati parte al Protocollo Addizionale (I) alle Convenzioni di Ginevra del 1977, ai sensi dell'articolo 36, sono obbligati nello studio, sviluppo, acquisizione o adozione di nuove armi, a determinare se il loro impiego possa essere proibito dal diritto internazionale a essi applicabile.

La proporzionalità richiede una ponderazione contestuale di due fattori: la possibilità di uccidere o ferire civili e danneggiare beni civili, da un lato, e il potenziale vantaggio militare dell'attacco, dall'altro. La determinazione del danno potenziale per i civili e i loro beni è determinabile con un certo livello di accuratezza, di conseguenza, i comandanti utilizzano già simulatori di danni collaterali per garantire che gli attacchi siano proporzionali rispettando, inoltre, il principio di precauzione (Schmitt, 2013). La questione se un sistema autonomo possa effettuare un'analisi di proporzionalità può essere ancora più gravosa di quella relativa alla capacità dei LAWS di rispettare il principio di distinzione. L'analisi di proporzionalità risulta particolarmente condizionata da fattori specifici e spesso in rapido mutamento per consentire di ridurla, ad esempio, a una regola secondo cui è possibile uccidere una vittima civile per ogni combattente; o due vittime civili per comandante di unità; o tre vittime civili per un carro armato distrutto. Inoltre, è fondamentale valutare se esistano mezzi e metodi di combattimento diversi in grado di ottenere il medesimo vantaggio militare ma

limitando i danni a civili e beni civili. Poiché la valutazione della proporzionalità richiede un peso relativo da attribuire agli interessi concorrenti, questi sistemi d'arma dovrebbero essere in grado di anticipare l'effetto di tutte le potenziali decisioni e il numero potenzialmente risultante di vittime civili, dovrebbero anche reagire alle mutevoli circostanze, essendo in grado di calcolare il vantaggio militare e determinare se il danno collaterale è accettabile.

Più complesso risulta invece il quadro dal punto di vista del diritto penale internazionale. Secondo lo Statuto della Corte Penale Internazionale (CPI), lanciare intenzionalmente un attacco, consapevoli che i danni a beni civili e ai civili siano eccessivi rispetto al vantaggio militare concreto e diretto conseguito, rappresenta un crimine di guerra perseguibile dalla Corte (art. 8 c. 2). L'elemento centrale nel perseguire una persona è rappresentato dal *mens rea* o elemento psicologico. Infatti, secondo quanto disposto dall'articolo 30 dello Statuto della CPI, una persona è penalmente responsabile solo se oltre l'elemento materiale – la condotta contraria allo Statuto – vi sia anche intenzione e consapevolezza. Se, almeno fino a oggi, risulta impensabile perseguire una macchina per una condotta criminale da essa presa autonomamente, è pur vero che la mancanza di un adeguato controllo e la potenziale imprevedibilità dei sistemi d'arma completamente autonomi rende difficile perseguire gli individui responsabili della programmazione e del dispiegamento dell'arma. Infatti, la capacità di elaborazione e apprendimento della macchina potrebbe superare le previsioni stesse di coloro i quali hanno programmato e dispiegato tali armi non rendendoli di fatto responsabili della condotta a causa dell'imprevedibilità del comportamento del sistema d'arma (Iaria, 2018).

6.4 Il ruolo dei consulenti giuridici e politici

Gli Stati, ratificando le Convenzioni di Ginevra e i suoi protocolli, si impegnano a rispettare e garantire il rispetto del diritto internazionale umanitario. La conoscenza del diritto applicabile durante le operazioni militari diventa dunque cruciale per rispettare e far rispettare gli obblighi esistenti. Nel rispetto di questi obblighi, l'articolo 82 del Protocollo Addizionale (I) stabilisce che «le Alte Parti contraenti in ogni tempo, e le Parti in conflitto in periodo di conflitto armato garantiranno che dei consiglieri giuridici siano disponibili, quando occorra, per consigliare i comandanti militari di livello appropriato circa l'applicazione delle Convenzioni e del presente Protocollo, e circa l'insegnamento appropriato da impartire in materia alle forze armate». In tempo di guerra il compito del consigliere giuridico è dunque di dare un parere sulle operazioni militari pianificate e in corso, verificare l'osservanza del diritto esistente nonché sulle implicazioni giuridiche legate alla condotta intrapresa dal comandante delle operazioni. In questo flusso di informazioni inviate e ricevute risultano fondamentali i canali di

trasmissione delle informazioni e la capacità dell'agente di adeguare la propria condotta al mutare del contesto operativo analizzato dal consigliere giuridico e dall'ordine impartito dal comandante delle operazioni.

Nell'arco degli anni accanto alla figura del consigliere giuridico, le Forze Armate hanno sviluppato consulenti specializzati in affari politici, culturali, religiosi e di genere. Tali figure, pur non espressamente previste dal punto di vista giuridico nelle condotte delle operazioni, risultano fondamentali. Per esempio, la figura del consulente politico è cruciale per il rispetto delle norme previste dal diritto internazionale umanitario. Nell'applicazione del principio di proporzionalità nell'attacco, per esempio, la valutazione relativa al vantaggio militare concreto e diretto è prerogativa del comandante in capo all'operazione supportato dai consulenti legali e politici incaricati di illustrare il piano giuridico nonché gli aspetti politici, economici e sociali che completano il quadro in cui l'azione militare viene svolta.

È chiaro dunque che il ruolo di queste due figure è fondamentale nella valutazione giuridica e politica che il comandante delle operazioni compie e negli ordini impartiti a tutti i livelli della catena di comando. Nel caso di sistemi d'arma autonomi, così come nel caso di esseri umani dispiegati in operazioni militari, vi è la necessità di garantire la responsabilità del comandante delle operazioni rispetto all'ordine dato. Se volessimo anche considerare i LAWS come agenti di diritto, ipotesi rigettata in precedenza, essi sarebbero posti alla fine di una catena di comando in cui il comandante delle operazioni è posto al vertice ed è coadiuvato nelle sue scelte dai consiglieri giuridici e politici. Pertanto, una diversa valutazione del contesto e del mutato quadro operativo deve essere prima vagliato da chi è responsabile dell'operazione e, se ne fa richiesta, dai consiglieri giuridici e politici. L'analisi della catena di comando e le valutazioni sulla responsabilità dell'operazione che ne conseguono, sembrano ancora una volta propendere verso la necessità che i sistemi d'arma autonomi siano subordinati ad un significativo controllo da parte degli esseri umani (Iaria, 2018).

6.5 L'umanità e la pubblica coscienza

La definizione della parola «umanità» rappresenta un esercizio piuttosto complesso; dal punto di vista etimologico deriva dal latino *humanĭtas-atis*, derivato di *humanus* «umano»; nel significato di «genere umano» (Dizionario Treccani). Ancor più difficile è definire l'umanità su un piano giuridico probabilmente perché le implicazioni di natura etica nel definire cosa rientri nel genere umano mutano al mutare del contesto storico, sociale e culturale. Forse più che di definizioni abbiamo bisogno di esempi per comprendere cosa possa significare applicare il principio di umanità in guerra (Iaria, 2018).

Così il giornalista e scrittore G. H. Perris descrisse la tregua di Natale del 1914 che

interessò il fronte delle Fiandre dove si confrontavano soldati britannici e tedeschi durante la prima guerra mondiale.

> "L'oscurità calò verso le 7 della vigilia di Natale, e con essa una calma improvvisa. I cecchini tedeschi sembravano essere scomparsi. Poi il suono del canto dei caroselli salì dalle trincee; e, a quel punto, i cecchini britannici smisero di puntare i loro fucili contro il nemico. Il coro magico affondò nelle tenebre e nuovamente riprese verso il cielo nero. Alcuni soldati britannici abbozzarono un applauso. "Urla dai tedeschi: 'Tu inglese, perché non vieni fuori?'" - così scrisse un ufficiale della R.F.A.- "e i nostri fieri fanti risposero, urlando, di 'Aspettare'." Alla fine uscirono; e, molto presto, fuochi e candele ardevano lungo i parapetti fino a quel momento vigilati incessantemente, mentre gli uomini fraternizzavano in mezzo a una folla, scambiandosi doni e concordando che la tregua durasse fino alla mezzanotte del giorno di Natale. "Tutto venne organizzato privatamente e avviato da uno dei nostri compagni. Difficile immaginarlo. L'unica cosa proibita era di apportare migliorie alle trincee.
>
> Se per un qualsiasi inconveniente fosse stato sparato un solo colpo, non sarebbe stato preso come un atto di guerra, e le scuse sarebbero state accettate; inoltre, la ripresa delle ostilità non sarebbe avvenuta senza il dovuto preavviso da entrambe le parti" (Perris, 1915, p.391).

In quell'occasione i soldati decisero spontaneamente di stabilire delle tregue sul fronte durante il periodo natalizio.

Se la notte di Natale del 1914 sul fronte fossero stati schierati sistemi d'arma autonomi non vi sarebbe stata alcuna tregua, sarebbe venuto meno quel principio di umanità tra esseri umani che, condividendo una fede religiosa, decisero in modo spontaneo di dichiarare una tregua. D'altra parte l'umanità è costellata da esempi di umanità, dove l'utilizzo ripetuto della stessa parola non è un refuso quanto piuttosto un'occasione per evidenziare il legame indissolubile tra l'uomo e la sua natura umana (Iaria, 2018).

Nel diritto internazionale umanitario le nozioni di umanità e pubblica coscienza sono state introdotte da quella che è nota come Clausola Martens. Durante la conferenza diplomatica in cui si discusse il Regolamento dell'Aja del 1899, il diplomatico prussiano Martens, chiese che venisse introdotto nel preambolo un paragrafo secondo cui: «i civili e i combattenti rimangono sotto la protezione e l'imperio dei principi del diritto delle genti quali risultano dalle consuetudini stabilite, dai principi di umanità e dai dettami della pubblica coscienza». La Clausola Martens è considerata una norma di diritto consuetudinario ed è stata inclusa anche nel Protocollo Addizionale (I) alle Convenzioni di Ginevra all'articolo 1.2 e nel preambolo del Protocollo Addizionale relativo ai conflitti armati non internazionali. La *ratio* della clausola nasce dal prevenire un'errata interpretazione secondo cui ciò che non è esplicitamente proibito dai trattati, sia considerato lecito e con questo spirito la Clausola Martens va vista, utilizzando il

linguaggio della CIG, come una rete di sicurezza per l'umanità (ICJ, *Legality of the threat or use of nuclearweapons, Advisory Opinion of 8 July 1996*). È parere di chi scrive che l'impossibilità di implementare il principio di umanità a sistemi d'arma completamente autonomi impedisce il rispetto della cornice giuridica esistente a queste tipologie di armi.

L'avanzare del diritto internazionale umanitario si è sempre caratterizzato dal ruolo svolto dalla società civile, accademici, leader politici, dagli Stati e dal Comitato Internazionale di Croce Rossa nel definire la coscienza pubblica. È un lavoro instancabile che ha trovato riconoscimento nel preambolo di diverse convenzioni approvate grazie alla competenza, e spesso intransigenza, della società civile affinché la coscienza pubblica prendesse la forma di obbligo internazionale. Nella recente approvazione del Trattato per la Proibizione delle Armi Nucleari, avvenuta nel 2017, l'ultimo paragrafo del preambolo sottolinea «il ruolo della coscienza pubblica nell'avanzamento dei princìpi dell'umanità come dimostrato dalla richiesta di eliminazione totale delle armi nucleari e riconoscendo gli sforzi a tal fine intrapresi dalle Nazioni Unite, dal Movimento internazionale della Croce Rossa e dal Mezzaluna Rossa, da altre organizzazioni internazionali e regionali, organizzazioni non governative, leader religiosi, parlamentari, accademici e dagli *hibakusha*».

Le obiezioni sollevate da diversi Stati, dal Movimento Internazionale di Croce Rossa e Mezzaluna Rossa, da accademici, scienziati, esperti e membri della società civile nonché da diversi leader politici e civici, così come la creazione della campagna per mettere al bando i robot killer, lasciano intendere come l'attenzione verso lo sviluppo e l'utilizzo di sistemi d'arma completamente autonomi sia alto e starà a questi attori internazionali definire la coscienza pubblica su un tema così controverso.

6.6 La posizione UE e italiana

In ambito UE, solo il Parlamento Europeo si è espresso nettamente contro le armi autonome. In una risoluzione adottata dal Parlamento il 27 febbraio 2014 dedicata ai droni, è stato inserito un punto relativo alle armi autonome, dove si chiede all'Alto Rappresentante dell'Unione per la politica estera e di sicurezza comune, agli Stati Membri e al Consiglio di "vietare lo sviluppo, la produzione e l'impiego di armi completamente autonome che consentono di sferrare attacchi senza alcun intervento umano".

In ambito UE, è tuttavia il Consiglio, operante nel quadro della politica estera e di sicurezza comune, ad essere competente ad esprimere la posizione dell'Unione in relazione al dibattito internazionale per lo sviluppo di una normativa inerente i LAWS.

Alla vigilia della quinta conferenza di riesame della Convenzione sulle CCW, tenutasi a dicembre 2016, il Consiglio dei Ministri degli Affari Esteri (14-15 novembre 2016) ha

apprezzato il lavoro del gruppo di esperti governativi in materia LAWS, evidenziando che allo sviluppo dei sistemi di armi autonome si applica il corpus di norme e principi del diritto internazionale umanitario e le altre regole rilevanti del diritto internazionale. Preme comunque sottolineare come il Consiglio, a differenza del Parlamento, non si sia espresso a favore di un divieto assoluto, quanto per una regolamentazione delle armi autonome (Ronzitti, 2018), consolidando e irrobustendo le norme di diritto umanitario già applicabili attraverso la definizione di principi operativi e linee guida di dettaglio.

L'ex Alto Rappresentante dell'UE Federica Mogherini, in un discorso al Parlamento Europeo l'11 settembre 2018, ha conseguentemente messo in rilievo che "Il diritto internazionale, compreso il diritto internazionale umanitario e il diritto dei diritti umani, si applica a tutti i sistemi di armi; gli esseri umani devono prendere decisioni in merito all'uso della forza letale, esercitare il controllo sui sistemi di armi letali che usano e rimanere responsabili delle decisioni sulla vita e sulla morte; la Convenzione delle Nazioni Unite su alcune armi convenzionali è il quadro adeguato per discutere di regolamentare questo tipo di armi; e, dato il duplice uso delle tecnologie emergenti, le misure politiche non dovrebbero ostacolare la ricerca civile, compresa l'intelligenza artificiale (AI)"(EEAS, 2018).

Nel quadro dei lavori del gruppo di esperti governativi della CCW, l'Italia ha contribuito al dibattito normativo, negoziando con gli altri partner europei i contenuti degli interventi UE ed intervenendo a titolo nazionale. Ad esempio, nell'aprile del 2018, il Rappresentante Permanente italiano alla Conferenza sul Disarmo, Amb. Gianfranco Incarnato, ha anch'egli ribadito che il corpus normativo esistente di diritto internazionale umanitario si applica pienamente a tutti i sistemi d'arma, inclusi quelli relativi ai potenziali LAWS ed ha pertanto messo in rilievo la necessità di un controllo umano rilevante su sistemi d'arma letali, anche ai fini della responsabilità in caso di violazioni. In linea con la posizione europea, l'Italia ha enfatizzato la rilevanza di una piena applicazione dell'art. 36 del Protocollo Addizionale alle Convenzioni di Ginevra del 1949. Come già descritto, l'art. 36 impegna gli Stati ad una verifica ex ante dei sistemi d'arma nella fase iniziale del *procurement*, potendosi ab origine verificare la compatibilità con il diritto internazionale umanitario di un sistema d'arma in progettazione. In Italia, la verifica che ogni sistema d'arma nella fase di studio, sviluppo, acquisizione o adozione rispetti il diritto internazionale umanitario è contemplata dal Codice dell'Ordinamento Militare del 2010Allo stadio attuale il processo di verifica si declina principalmente in un controllo di carattere parlamentare, condotto dal Ministero della Difesa in cooperazione con le Commissioni Difesa della Camera e del Senato, che hanno il compito di vagliare e deliberare qualsiasi spesa relativa allo studio, sviluppo, acquisizione e adozione di nuovi sistemi d'arma. In sede di CCW si è pertanto potuto positivamente mettere in luce come in Italia vi sia già un robusto meccanismo di controllo parlamentare sulla fase di *procurement* dei sistemi

d'arma. Tale meccanismo potrebbe tuttavia essere ulteriormente rafforzato sulla base delle raccomandazioni e migliori pratiche, che emergeranno dai lavori del gruppo di esperti governativi della Convenzione in tema di concreta attuazione dell'art. 36.

Inoltre, nel dibattito sulla caratterizzazione dei LAWS, l'Italia ha ribadito come il lavoro del GGE dovesse focalizzarsi su quei sistemi d'arma letali che sono pienamente autonomi, ribadendo che tali sistemi ancora non esistono, ma che possono essere sviluppati in futuro ponendo un problema di compatibilità con le regole del diritto internazionale umanitario (GGE, Characterization of LAWS, 2018).

In un recente intervento durante i lavori della Giornata di studio su "Intelligenza artificiale, sicurezza, responsabilità, etica" promossa dal Segretariato generale della Difesa e Direzione nazionale degli armamenti in collaborazione con il Pontificio consiglio della Cultura della Santa Sede, il Ministro della Difesa Guerini ha affermato che parlando di intelligenza artificiale che ormai domina molte delle nostre attività quotidiane, è di vitale importanza "definire in modo chiaro e condiviso i limiti e le condizioni di autonomia di tali macchine" (Analisi Difesa, 2019). In particolar modo quando viene considerato il settore militare, il Ministro ha affermato che "sappiamo benissimo che sistemi di armamento autonomi già esistono[…], ma è evidente che in campo militare esistono significative implicazioni etiche e legali" (Analisi Difesa,2019), pertanto esiste la concreta necessità di "individuare un adeguato sistema giuridico entro il quale poter collocare la robotica autonoma"(Analisi Difesa, 2019) in accordo con quanto già sancito dalla Commissione Giuridica del Parlamento Europeo.

Nel corso dell'ultimo meeting del Gruppo di Esperti Governativi sui LAWS, tenutosi tra il 20 e il 21 agosto 2019, dopo una lunga fase di negoziazione, si è finalmente trovato il consenso su un rapporto finale, che include la definizione di 11 principi guida nel campo dello sviluppo delle LAWS condivisi da tutti gli Stati partecipanti. Tali principi stabiliscono che:

(a) Il diritto internazionale umanitario continua ad applicarsi pienamente a tutti i sistemi di armamento, compreso il potenziale sviluppo e utilizzo di sistemi letali di armi autonome.
(b) La responsabilità umana per le decisioni sull'uso dei sistemi d'arma deve essere mantenuta e che la responsabilità non può essere trasferita alle macchine. Questo dovrebbe essere considerato durante l'intero ciclo di vita dei sistemi d'armi.
(c) L'interazione uomo-macchina, che può assumere varie forme ed essere implementata in varie fasi del ciclo di vita di un'arma, dovrebbe garantire che il potenziale utilizzo di sistemi d'arma basati su tecnologie in grado di produrre sistemi di armi autonome letali siano conformi alle norme internazionali applicabili, e in particolare al diritto internazionale umanitario. Nel determinare la qualità e l'estensione dell'interazione uomo-macchina, una serie di fattori dovrebbe essere considerata includendo il contesto operativo e le caratteristiche e le capacità delle armi nel loro complesso.

(d) La responsabilità per lo sviluppo, il dispiegamento e l'utilizzo di qualsiasi sistema di armi emergenti nel quadro della CCW deve essere garantito in conformità al diritto internazionale applicabile, anche attraverso l'utilizzo di tali sistemi all'interno di una catena responsabile di comando e controlli umani.
(e) In conformità con gli obblighi degli Stati e in base al diritto internazionale, nello studio, nello sviluppo, acquisizione o adozione di una nuova arma, mezzi o metodi di combattimento, è necessario stabilire se il suo impiego, in alcune o in tutte le circostanze, sia vietata dal diritto internazionale.
(f) Quando si sviluppano o acquisiscono nuovi sistemi di armi basati su tecnologie emergenti nell'area dei LAWS, dovrebbero essere considerati dei sistemi di sicurezza fisica, adeguate garanzie non fisiche (compresa la sicurezza informatica contro l'*hacking* o lo *spoofing* dei dati), il rischio di acquisizione da parte di gruppi terroristici e il rischio di proliferazione di tali sistemi.
(g) Le valutazioni dei rischi e le misure di mitigazione dovrebbero far parte della progettazione, dello sviluppo, delle prove e ciclo di spiegamento delle tecnologie emergenti in tutti i sistemi d'arma.
(h) Dovrebbe essere preso in considerazione l'uso di tecnologie emergenti nell'area dei LAWS nel rispetto e in conformità del diritto internazionale umanitario e delle altre leggi internazionali applicabili.
(i) Nell'elaborare potenziali misure politiche, tecnologie emergenti nell'area dei LAWS, queste non dovrebbero essere antropomorfizzate.
(j) Le discussioni e le eventuali misure politiche prese nel contesto della CCW non dovrebbero ostacolare il progresso o l'accesso a usi pacifici di tecnologie autonome intelligenti.
(k) La CCW offre un quadro adeguato per affrontare il problema delle tecnologie emergenti nel quadro dei LAWS nel rispetto degli obiettivi e delle finalità del Convenzione, che cerca di trovare un equilibrio tra necessità militare e considerazioni umanitarie.

La Conferenza degli Stati parte alla CCW riunitasi dal 13 al 15 novembre 2019, con l'elaborazione del report finale, ha deciso di supportare i principi guida proposti dalla GGE e di convocare il Gruppo dal 22 al 26 giugno 2020 e dal 10 al 14 agosto 2020 e 10-20 giorni nel 2021 per proseguire nel lavoro con lo scopo di considerare:

- i principi guida per un ulteriore sviluppo ed elaborazione;
- il lavoro sugli aspetti legali, militari e tecnologici;
- le conclusioni del Gruppo.

I prossimi lavori del GGE si focalizzeranno sull'obiettivo di rendere i principi guida quanto più operativi possibili in modo tale da essere concretamente implementati attraverso politiche e pratiche a livello nazionale, anche in connessione con raccomandazioni puntuali per attuare il disposto dell'art 36 del Protocollo Addizionale

alle Convenzioni di Ginevra del 1949 (richiamato dai principi guida alla lettera e).

Merita rilevare come lungo tutto il corso dei lavori del Gruppo di Esperti Governativi della Convenzione su Certe Armi Convenzionali, i partner europei e l'Italia hanno lavorato intensamente per promuovere un approccio costruttivo e condiviso nell'analisi delle implicazioni di diritto umanitario poste dalle LAWS, portando avanti una linea di pensiero incentrata a questo stadio sullo sviluppo di strumenti normativi non legalmente vincolanti, come la definizione dei sopramenzionati principi guida. Tale approccio, frutto di un proficuo dialogo avviato da alcuni Stati dell'Unione Europea e su cui poi si è allineata l'intera Unione, risulta particolarmente interessante e a tratti innovativo nel contesto della CCW. Infatti, la scelta di principi guida non giuridicamente vincolanti, ma comunque efficaci per costruire un iniziale *consensus* intorno a un tema così delicato e ricco di sviluppi tecnologici nel breve periodo sul crinale civile-militare (dual use), offre un indubbio beneficio nel muoversi con gradualità laddove le attuali condizioni politiche rischiano di far arenare ogni processo normativo.

D'altra parte, il diritto internazionale umanitario, a oltre settanta anni dalla firma delle Convenzioni di Ginevra, vede oggi una certa resistenza da parte di alcuni Stati nel giungere a nuovi strumenti giuridicamente vincolanti o comunque rafforzare e implementare le misure già esistenti. Ne è un esempio l'impossibilità di trovare un *consensus* tra gli Stati per trasmettere un report sul rafforzamento del rispetto del diritto internazionale umanitario alla 33^ Conferenza Internazionale di Croce Rossa e Mezzaluna Rossa dopo il meeting del 3-5 dicembre 2018. Per questo, la scelta di procedere attraverso dichiarazioni politiche sta mostrando tutta la sua efficacia.

Nel contesto del GGE sui LAWS, la linea di pensiero adottato dall'Unione Europea ha portato ad un risultato più che lusinghiero con la definizione degli undici principi guida che offrono un primo quadro (*framework*) condiviso su cui lavorare e confermano la centralità della CCW nel processo normativo inerente i LAWS.

Originariamente promossi da un *Food for Thought Paper* presentato da Belgio, Irlanda e Lussemburgo, i principi guida andrebbero ulteriormente posti in rilievo attraverso una dichiarazione politica ampiamente condivisa da adottare ad esempio in occasione della prossima Conferenza degli Stati parte, come auspicato da molti partner dell'Unione Europea e dall'Italia.

Altri Stati invece auspicano che dai principi guida si possa aprire fin da subito un percorso per giungere a framework giuridicamente vincolanti.

6.7 Conclusioni

Il CEO di Apple, Tim Cook, durante la cerimonia in onore dei laureati nel 2017 del MIT di Boston, ha ben spiegato il ruolo della tecnologia di fronte alle sfide poste dall'intelligenza artificiale.

> La tecnologia può fare grandi cose. Ma non vuole fare grandi cose. Non vuole fare niente. Questo ruolo spetta a noi. Spetta ai nostri valori e al nostro impegno verso i nostri familiari, ai nostri amici, le nostre comunità, spetta al nostro amore per la bellezza e alla convinzione che le nostre vite sono interconnesse, al nostro senso civico e alla nostra gentilezza. Non sono preoccupato dell'intelligenza artificiale che dà ai computer la capacità di pensare come gli umani. Sono più preoccupato per le persone che pensano come computer senza valori o compassione, senza preoccuparsi delle conseguenze. Questo è ciò di cui abbiamo bisogno per proteggerci. Perché se la scienza è una ricerca nell'oscurità, allora le scienze umane sono una candela che ci mostra dove siamo stati e il pericolo che ci aspetta.

Lo sviluppo di nuove tecnologie spesso apre dibattiti di natura etica e giuridica sulle implicazioni delle stesse. Intelligenza artificiale, *deep learning* e algoritmi complessi sono temi che fino ad oggi hanno visto i colossi della tecnologia – Google, Facebook, Apple – pionieri in questo settore, non tenendo conto di cosa significhi applicare queste tecnologie al campo militare, sottovalutando, in ultima istanza, il rischio concreto di delegare alla "coscienza" di una macchina il diritto di una persona di vivere o morire. Sarà dunque necessaria una maggior sinergia tra aziende dell'industria tecnologica, Stati e aziende militari per valutare attentamente sia le opportunità che i rischi insiti nel sviluppare sistemi di AI sempre più complessi.

Se è vero che il progresso della scienza è inevitabile, è pur vero che applicare i progressi scientifici al campo militare non è inevitabile, piuttosto è una scelta. E allora questa scelta non può non tenere conto delle esigenze di umanità dei conflitti armati e del progredire del diritto internazionale che ha previsto la repressione dei crimini internazionali attraverso la Corte Penale Internazionale. La responsabilità nella condotta militare rappresenta un punto centrale al fine di rispettare e garantire il rispetto del diritto dei conflitti armati. Questo principio non può essere oggetto di ripensamenti a causa di sistemi d'arma che, per il loro elevato livello di autonomia e quindi di imprevedibilità, non permetterebbero di accertare l'eventuale responsabilità penale della condotta. L'elemento psicologico, il *mens rea*, non può che essere un elemento tipico dell'essere umano ed è dunque giusto relegare i LAWS all'alveo delle macchine piuttosto che degli agenti, condividendo quanto sostenuto da Sassòli che solo gli esseri umani sono destinatari del diritto internazionale umanitario (Iaria, 2018).

Bibliografia

Analisi Difesa (2019).“Guerini: perplessità sull’utilizzo spinto dell’intelligenza artificiale”. In *Analisi Difesa*, 25 settembre, https://www.analisidifesa.it/2019/09/guerini-perplessita-sullutilizzo-spinto-dellintelligenza-artificiale/

De Vitoria F. (1539). *De jure belli*. Trad. it. Roma: Laterza, 2005.

Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems Geneva, 25–29 March 2019 and 20–21 August 2019 Agenda Item 5, Focus of work of the Group of Governmental Experts in 2019, *Food for Thought Paper submitted by Belgium, Ireland and Luxembourg*.https://www.unog.ch/80256EDD006B8954/(httpAssets)/F3A850205D48D5A4C125841E00351CD7/$file/CCW_GGE.1_2019_WP.4.pdf

European External Action Service, *Speech by High Representative/Vice-President Federica Mogherini at the plenary session of the European Parliament on Autonomous Weapons Systems,* Strasbourg, 11 September 2018. https://eeas.europa.eu/headquarters/headquarters-homepage/50307/node/50307_et

Gentili A (2008). *Il diritto di guerra (De jure belli libri cap. VI del libro I, 1598)*. Roma: Giuffrè-

Grotius H. (1625). *De jure belli acpacis*. Trad. It. Firenze: Nabu Press, 2011.

Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects (2019).*Draft Report of the 2019 session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (Annex 4).*21 August, https://www.unog.ch/80256EDD006B8954/(httpAssets)/5497DF9B01E5D9CFC125845E00308E44/$file/CCW_GGE.1_2019_CRP.1_Rev2.pdf

Iaria A. (2018). “Da autonomi a completamente autonomi: l’applicazione dell’intelligenza artificiale nei sistemi d’arma autonomi (LAWS)”. In *Rassegna della Giustizia Militare*, 6: 24-32, https://www.difesa.it/Giustizia_Militare/rassegna/Bime-strale/2018/Documents/RGM_6_2018.pdf

ICJ (1996). *Legality of the threat or use of nuclear weapons, Advisory Opinion of 8 July 1996,* https://www.icj-cij.org/ en/case/93

ICRC (1987). *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949*.MartinusNijhoff Publishers.

ICRC (1995). *Vienna Diplomatic Conference Achieves New Prohibition on Blinding Laser Weapons and Deadlock on Landmines*, 13 October.

ICRC (2018). *Ethics and autonomous weapon systems: An ethical basis for human control?*, 3 aprile, https://www.icrc. org/en/document/ethics-and-autonomous-weapon-systems-ethical-basis-human-control

Messineo A. (1971). "Il diritto umanitario". In *La civiltà cattolica*, 2896: 319-331.

Moyes R. (2016). "Meaningful human control over individual attacks". In ICRC, *Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons*. Switzerland: Versoix, 15-16 March: 46-52.

CCW Group of Governmental Experts on Lethal Autonomous Weapons Systems Geneva (9 – 13 April 2018). *Statement by Ambassador Gianfranco Incarnato, Permanent Representative of Italy to the Conference on Disarmament, Characterization of LAWS*. https://www.unog.ch/80256 EDD 006B8954/(httpAssets)/36335330158B5746C1258273003903F0/$file/2018_LAWS6a_Italy.pdf

CCW Group of Governmental Experts on Lethal Autonomous Weapons Systems Geneva (9 – 13 April 2018). *Statement by Ambassador Gianfranco Incarnato, Permanent Representative of Italy to the Conference on Disarmament, General exchange of views.* https://www.unog.ch/80256 EDD006B8954/(httpAssets)/8F7600027923383BC12582730038E522/$file/2018_LAWS GeneralExchange_Italy.pdf

Perris G.H. (1915). *The campaign of 1914 in France and Belgium.* Henry Holt and Company.

Ronzitti N. (2018). "Uso e sviluppo delle armi autonome. Prospettive per un controllo a livello internazionale*".* In *Studi per il Parlamento.* Roma: Camera: 12.

Rousseau J. J. (1762). *Il contratto sociale*. Tr. it. Milano: Istituto Editoriale Moderno, 2016: 19.

Schmitt M. (2013). *Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics*, Harvard National Security Journal Feature.

Schuller A. (2016). "Focusing the debate on autonomous weapon systems: A new approach to linking technology and IHL". In ICRC, *Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons.*, Switzerland: Versoix, 15-16 March.

SunTzu (2013). *L'arte della guerra*. Tr. it. Milano: Feltrinelli (ed. or. VI sec. a.C).

Cap. 7
Il dibattito in seno alla Convenzione su certe armi convenzionali CCW A Ginevra

7.1 La comunità scientifica e i problemi etici sollevati dalle armi autonome

La comunità dei ricercatori dell'IA e della robotica ha contribuito con varie iniziative a sensibilizzare il mondo politico e l'opinione pubblica sui problemi etici posti dallo sviluppo dei sistemi d'arma autonomi (vedi cap. 1). Ricordiamo qui solo alcune iniziative tra le più salienti.

Nel 2012, 270 ricercatori in informatica e robotica, provenienti da 37 paesi diversi, sottoscrissero un appello pubblico con la richiesta di proibire le armi autonome[92]. L'appello, promosso e pubblicato dall'ICRAC (*International Committee for Robot Arms Control*), pone l'accento sulle questioni di responsabilità morale e giuridica sollevate dalle armi autonome, sottolineando la possibilità di un vuoto di responsabilità che potrebbe emergere allorché si sposta dall'essere umano alla macchina la decisione di impiegare la forza in un'azione bellica: non vi sarebbe più un *qualcuno* al quale chiedere conto di eventuali esiti illegali o moralmente inaccettabili di una tale decisione automatica, poiché essa è stata operativamente presa da un *qualcosa* che non ha i titoli per poter essere biasimato, punito o in ogni caso considerato responsabile dell'accaduto. In altri termini, l'appello degli scienziati motiva la richiesta di proibire le armi autonome principalmente sull'obbligo morale di collegare ogni atto di forza compiuto nel corso di un conflitto bellico alle decisioni e all'assunzione di responsabilità da parte di uno o più esseri umani.

Nel 2013, un nutrito gruppo di organizzazioni non governative lanciò la campagna SKR (*Stop Killer Robots*), con l'obiettivo di promuovere l'adozione di un trattato che proibisca lo sviluppo, la produzione, il possesso e l'uso delle armi autonome[93]. Tra i promotori della campagna figurano associazioni e gruppi formati da ricercatori che operano in vari settori scientifici, come l'organizzazione delle *Pugwash Conferences on Science & World Affairs*, insignita del Premio Nobel per la pace nel 1995 e il già ricordato ICRAC (vedi cap. 8).

Nel 2015, una lettera aperta per la messa al bando delle armi autonome fu presentata all'avvio dei lavori del principale congresso mondiale nel settore dell'IA (l'IJCAI – *International Joint Conference on Artificial Intelligence*). Migliaia di ricercatori dell'IA e della robotica firmarono questa lettera aperta[94], chiedendo di proibire definitivamente lo sviluppo, la produzione, la vendita e l'uso delle armi autonome. La

[92]https://www.icrac.net/the-scientists-call/.
[93]https://www.stopkillerrobots.org/ Ad essa aderiscono anche IRIAD e Uspid.
[94]http://futureoflife.org/open-letter-autonomous-weapons/.

lettera aperta esorta i ricercatori dell'IA e della robotica a seguire le orme delle comunità dei biologi e dei chimici, i quali contribuirono attivamente alle iniziative che sfociarono nella messa al bando delle armi biologiche e chimiche; e a seguire le orme della comunità dei fisici, i quali contribuirono attivamente alle iniziative che sfociarono nella messa al bando dei laser accecanti**.**

La lettera aperta si concentra sugli effetti negativi che lo sviluppo e l'uso delle armi autonome potrebbero avere per la sicurezza dei popoli, la pace e la stabilità internazionale: "Una questione centrale che l'umanità deve affrontare oggi riguarda l'avvio oppure la prevenzione di una corsa globale alle armi dell'IA. Se una delle maggiori potenze militari imboccherà la strada dello sviluppo delle armi autonome, una nuova corsa globale alle armi sarà praticamente inevitabile. E il punto di arrivo di una tale traiettoria tecnologica è del tutto evidente: le armi autonome diventeranno i kalashnikov del domani. A differenza delle armi nucleari, le armi autonome non richiedono materie prime costose o difficili da acquisire; si diffonderanno ovunque; saranno prodotte in grandi quantità e a costi molto bassi; faranno la loro comparsa sul mercato nero delle armi, nelle mani dei terroristi, nelle mani di dittatori interessati al controllo della popolazione e di signori della guerra intenti a progettare operazioni di pulizia etnica. Le armi autonome sono uno strumento ideale per compiere esecuzioni politiche mirate, per destabilizzare nazioni intere, soggiogare popolazioni e sterminare selettivamente un gruppo etnico."

Queste ragioni morali per la messa al bando delle armi autonome sono principalmente riconducibili all'etica delle conseguenze. Appelli dello stesso tenore sono stati in seguito sottoscritti da imprenditori operanti nei settori dell'informatica, della robotica e dell'intelligenza artificiale[95], da gruppi di loro dipendenti, da ricercatori operanti in settori affini.

7.2 Le armi autonome alla *Convention on Certain Conventional Weapons*

Stimolato anche dalle iniziative prese da ampi settori della società civile, il confronto diplomatico e politico sulle armi autonome ha preso le mosse dall'esame delle nuove questioni giuridiche ed etiche sollevate da questa tipologia di armi convenzionali. Tra le tante prese di posizione al riguardo, ci limitiamo qui a ricordarne solo alcune nell'ambito dell'Unione Europea e dei suoi stati membri. Il 4 luglio 2018, il Governo Federale tedesco ha fornito la seguente risposta a una interrogazione parlamentare sulle armi autonome: "The Federal Government is active lyengaged in promoting a ban on lethal autonomous weapon systems lacking human control over the use of deadly force." Il 12 settembre 2018, il Parlamento Europeo ha approvato una risoluzione con la

[95]https://futureoflife.org/2017/08/20/killer-robots-worlds-top-ai-robotics-companies-urge-united-nations-ban-lethal-autonomous-weapons/.

quale si invitano "the [High Representative for Foreign Affairs and Security Policy], the Member States and the Council to work towards the start of international negotiations on a legally binding instrument prohibiting lethal autonomous weapon systems."

E con la mozione 1-01776 approvata nella seduta del 6 dicembre 2017, la Camera dei deputati impegna il Governo italiano "a continuare a partecipare attivamente al dibattito internazionale in corso in particolare nell'ambito della Convenzione sulle armi convenzionali, di concerto con i principali *partner* dell'Italia […] con l'obiettivo di arrivare a una […] regolamentazione internazionale dei sistemi d'arma di tipo *Lethal Autonomous Weapons System* (Laws) […]".

Come ricorda la mozione della Camera dei Deputati, il confronto internazionale sulle armi autonome si è principalmente svolto, a partire dal 2014, nella sede di Ginevra delle Nazioni Unite, nella cornice istituzionale della "Convenzione delle Nazioni Unite sulla proibizione o la limitazione dell'uso di alcune armi convenzionali che possono essere considerate eccessivamente dannose o aventi effetti indiscriminati" (più sinteticamente Convenzione sulle armi convenzionali o CCW, abbreviazione di *Convention on Certain Conventional Weapons*).

Le sessioni del CCW dedicate alle armi autonome inizialmente (2014-2016) hanno avuto scopi conoscitivi riguardo alle tecnologie abilitanti per le armi autonome e di approfondimento delle questioni etiche e legali. Nel 2017 è stato costituito un *Group of Governmental Experts* (GGE) in seno al CCW, con il mandato di vagliare la compatibilità delle armi autonome con le norme del diritto internazionale umanitario, e di avanzare su questa base delle raccomandazioni da sottoporre alla valutazione e alle decisioni degli Stati membri del CCW.

Il CCW si è dovuto all'inizio confrontare con il problema di delimitare il suo oggetto di indagine e di chiarire i termini delle questioni sul tappeto. Punto di partenza obbligato è stato un esame delle proprietà salienti delle armi autonome, richiamate già nella direttiva sull'autonomia dei sistemi d'arma pubblicata nel 2012 dal Dipartimento della Difesa degli Stati Uniti d'America (DoD): un'arma autonoma deve essere in grado, dopo la sua attivazione, "di selezionare e di attaccare gli obiettivi senza ulteriori interventi da parte di un operatore umano" (DoD, 2012). Anche il Comitato Internazionale della Croce Rossa (ICRC) ha fatto riferimento a caratteristiche analoghe: le armi autonome devono possedere "la capacità di selezionare e di attaccare gli obiettivi in modo indipendente; godere cioè di autonomia nelle 'funzioni critiche' di acquisire, tracciare, selezionare e ingaggiare un obiettivo" (ICRC, 2014).

Le caratteristiche salienti individuate dal DoD e dell'ICRC hanno fornito un buon punto di partenza per la discussione politica e diplomatica sulle armi autonome in seno al CCW, ma in tutta evidenza non sono sufficienti a delimitare con una certa precisione il campo di analisi e discussione per il CCW. Si potrebbe infatti ragionevolmente sostenere che anche una mina antiuomo soddisfa la condizione per l'autonomia posta dal DoD, poiché essa è capace, sia pure in modo piuttosto primitivo, di selezionare in

funzione del peso rilevato dai suoi sensori un obiettivo dopo essere stata attivata e di attaccarlo senza richiedere ulteriori interventi da parte di un operatore umano. In definitiva, l'impostazione al problema dell'autonomia adottata dal DoD e dall'ICRC non permette di distinguere adeguatamente l'autonomia operativa di armi costruite in base alle tecnologie più avanzate della robotica e dell'IA, le quali sollevano problemi etici nuovi e urgenti, dall'autonomia operativa di altri sistemi, come le mine antiuomo, che pongono questioni etiche già adeguatamente riconosciute e affrontate nella Convenzione per la proibizione dell'uso, stoccaggio, produzione, vendita di mine antiuomo e relativa distruzione (Trattato di Ottawa).

Per superare la difficoltà insita nell'impostazione al problema dell'autonomia adottata dal DoD e dall'ICRC, è utile considerare la proposta della ONG britannica *Article 36*, avanzata in un rapporto pubblicato nel 2013, nel quale si pone l'accento sulla necessità di assicurare che gli attacchi sferrati da *tutti* i sistemi d'arma siano soggetti a un "controllo umano significativo" (Article 36, 2013). Questa proposta ha avuto il merito di aver individuato per prima nell'elemento umano la questione chiave da affrontare nella riflessione sull'autonomia dei sistemi d'arma. La preoccupazione di un'interruzione della catena di responsabilità nelle azioni belliche ha motivato l'introduzione dell'aggettivo "significativo" nella frase "controllo umano significativo". Con questa qualifica, si intende escludere ogni forma di controllo puramente *nominale* di un sistema d'arma. Ciò potrebbe accadere se il controllore umano non ha a sua disposizione tempo sufficiente o informazioni adeguate per intervenire con cognizione di causa sul sistema d'arma.

La formula del controllo umano significativo è stata accolta con interesse da molti attori del dibattito al CCW. Kerstin Vignard, vice-direttrice dell'UNIDIR (United Nations Institute for Disarmament Research), ha osservato: "In definitiva, la questione dell'autonomia riguarda il tipo di controllo/vigilanza che riteniamo debba essere appannaggio degli esseri umani nell'impiego di strumenti per l'esercizio della violenza" (UNIDIR, 2016). Con una mozione presentata nel 2018 al CCW, Austria, Cile e Brasile hanno richiesto che il GGE proceda i suoi lavori con il mandato di negoziare uno strumento giuridico che sia vincolante per tutti gli Stati, finalizzato ad "assicurare un controllo umano significativo sulle funzioni critiche dei sistemi d'arma autonomi con effetti letali" (Austria *et al.*, 2018). Un interesse convergente della comunità internazionale per la formula del Controllo Umano Significativo (CUS) è emerso con la pubblicazione nel 2018 di un documento del GGE nel quale si afferma che "è necessario preservare la responsabilità umana per le decisioni che riguardano gli usi letali della forza" (GGE, 2018).

Varie caratteristiche della formula CUS spiegano perché essa sia stata accolta con tanto favore. L'idea di imporre un controllo umano significativo su qualsiasi sistema d'arma può essere afferrata facilmente, anche in assenza di conoscenze specifiche sulle tecnologie militari più avanzate. La formula CUS si avvantaggia della sua "ambiguità

costruttiva" (Crootof, 2016, pp. 58-60). Essa è ambigua poiché non esplicita in che cosa debba consistere un controllo veramente significativo sui sistemi d'arma; ma è costruttiva, perché esprime un'esigenza condivisa, all'intersezione di posizioni diplomatiche e politiche che differiscono tra loro per altri aspetti. Inoltre, la formula CUS consente di mettere da parte la questione definitoria "Che cos'è un'arma autonoma?", ponendo al centro della discussione una questione normativa: "Quale controllo umano si deve esercitare su ogni sistema d'arma?". Poiché la questione normativa riguarda ogni sistema d'arma, essa riguarda in particolare le armi che meritano di essere chiamate autonome, comunque sia definita la loro autonomia.

Lo spostamento di attenzione sulla questione normativa consente di accantonare le inconcludenti diatribe preliminari su cosa sia un'arma *veramente* autonoma, discussioni vaghe e malferme sulla distinzione tra sistema automatico e sistema autonomo, concezioni inadeguate – perché troppo inclusive o troppo restrittive – dell'autonomia operativa di un sistema tecnologico (Amoroso e Tamburrini, 2019).

7.3 Il controllo umano significativo e i suoi contenuti

Se la formula CUS offre un punto di partenza promettente per affrontare le questioni normative sollevate dalle armi autonome, una soluzione adeguata a tali questioni deve dare una risposta alla domanda: "Che cosa rende il controllo umano su un sistema d'arma veramente 'significativo'?" Per affrontare questa domanda, è utile richiamare i principi dell'etica dei doveri elencati e discussi nel capitolo 1:

(i) Rispetto delle norme del diritto internazionale umanitario nella condotta delle azioni belliche (principi di distinzione e proporzionalità).
(ii) Mantenimento della catena delle responsabilità nelle azioni belliche.
(iii) Rispetto della dignità degli esseri umani che possono subire l'uso della forza nelle azioni belliche.

Per garantire il rispetto di questi vincoli, è necessario assegnare all'essere umano alcuni ruoli fondamentali nel controllo dei sistemi d'arma (Amoroso e Tamburrini, 2019; Tamburrini, 2020):

1. Il controllo umano deve intervenire come sistema ausiliario di salvaguardia (*fail-safe*), per contribuire a impedire che il malfunzionamento o un comportamento imprevisto del sistema d'arma sfoci in un attacco diretto contro la popolazione civile e i suoi beni, in un danno collaterale eccessivo o in altre violazioni del diritto internazionale umanitario.
2. Il controllo umano deve funzionare come catalizzatore di responsabilità, assicurando la sussistenza delle condizioni per attribuire delle responsabilità personali nel caso di violazioni del diritto internazionale umanitario.

3. Il controllo umano deve garantire il rispetto della dignità degli esseri umani sottoposti a decisioni di vita o di morte in un conflitto armato, consentendo di ricondurre a un agente morale – piuttosto che a una macchina priva di tale caratteristica – le decisioni che hanno un impatto sulla loro vita, sulla loro integrità fisica e sui beni dei quali dispongono.

A proposito della prima condizione, è importante sottolineare il contributo che un essere umano può dare per prevenire i cosiddetti "disastri dell'IA" nelle azioni belliche. Un sistema dell'IA può indurre violazioni del diritto internazionale umanitario a causa di errori percettivi o di valutazione nei quali gli esseri umani non incorrono normalmente o che possono individuare utilizzando indizi contestuali difficilmente accessibili alla macchina (vedi capp. 1 e 2). L'importanza di questa funzione per un controllo umano veramente significativo può essere evidenziata ricordando il famoso caso del falso allarme nucleare nel quale fu coinvolto come protagonista il colonnello dell'aviazione sovietica Stanislav Yevgrafovich Petrov. La mattina del 26 settembre 1983, il sistema sovietico OKO di *early warning* per gli attacchi nucleari registrò il lancio in rapida sequenza di cinque missili balistici, partiti dal territorio degli Stati Uniti e diretti verso l'Unione Sovietica. Petrov, ufficiale responsabile di turno, arrivò a concludere che si trattava di un falso allarme dovuto a un malfunzionamento del sistema. In base a tale convincimento, Petrov contravvenne alla consegna di trasmettere ai superiori la notizia dell'allarme segnalato dal sistema, notificando invece un episodio di malfunzionamento. La diagnosi di Petrov si rivelò giusta. Ricevendo la segnalazione di lancio generata da OKO, il comando sovietico avrebbe potuto decidere di rispondere con un lancio di missili balistici, scatenando così una guerra nucleare devastante.

7.4 Verso un trattato internazionale sul controllo umano significativo delle armi autonome?

È importante a questo punto chiedersi se la richiesta di imporre un controllo umano significativo su ogni sistema d'arma implichi in modo necessario e uniforme l'eliminazione dell'autonomia operativa dei sistemi d'arma, per quanto riguarda specificamente le funzioni critiche di selezione e attacco di un obiettivo militare evidenziate dal DoD e da ICRC. Per affrontare questo interrogativo, osserviamo anzitutto che esistono molti sistemi d'arma i quali sono da tempo utilizzati dalle forze armate di vari stati e soddisfano le funzioni critiche individuate dal DoD e dall'ICRC, senza aver tuttavia suscitato obiezioni di incompatibilità con il diritto internazionale umanitario o altre obiezioni di tipo etico o giuridico.

Uno di questi sistemi d'arma è il sistema mobile antimissile *Iron Dome*, utilizzato in Israele per monitorare e neutralizzare con il lancio di missili intercettori i razzi ed altri

proiettili balistici diretti verso il territorio israeliano[96]. Gli operatori militari si limitano a posizionare sul terreno *Iron Dome* e a circoscrivere l'area che esso deve monitorare e proteggere. Dopo aver compiuto queste operazioni preliminari, gli operatori abilitano *Iron Dome* ad operare in piena autonomia (vedi cap. 3). Dotato di analoghe capacità di intercettazione, il sistema NBS (*Nächstbereichschutzsystem*) *MANTIS* è utilizzato dalle forze armate tedesche per proteggere i soldati e le installazioni militari da proiettili balistici in arrivo[97]. I tempi di reazione di *Iron Dome* e *MANTIS* – come pure del *Phalanx* e del *Centurion C-Ram* in dotazione alle forze armate statunitensi – sono inferiori a quelli di un operatore umano e consentono perciò di proteggere persone e cose con maggiore efficacia di un sistema che sia asservito a un operatore umano per la decisione ultima di attacco (Rossi, 2019, p. 10). Il loro uso come scudi protettivi di difesa non è stato mai messo in discussione da una prospettiva etica o giuridica. E tuttavia entrambi i sistemi soddisfano le proprietà salienti di selezione e attacco di un obiettivo individuate dal DoD e dall'ICRC.

Il CUS su questi sistemi difensivi può essere mantenuto senza privarli della loro autonomia nelle funzioni critiche di selezione e attacco di un obiettivo. È infatti sufficiente che un operatore umano si accerti periodicamente, in fase di pianificazione e in corso d'opera, che non vi siano oggetti o persone da proteggere nel campo di azione dell'arma. Ma questo stesso tipo di controllo non è sufficiente per garantire un controllo umano significativo su tutte le armi autonome esistenti. Vediamo perché, facendo riferimento a un altro esempio concreto, che riguarda il sistema israeliano *Harpy* (vedi cap. 3).

Le armi autonome già operative possono essere distinte, ai fini della discussione sul CUS, in armi difensive e offensive (Boulanin e Verbruggen, 2017, p. VII). Abbiamo visto che il sistema israeliano *Iron Dome* e il sistema tedesco *MANTIS* sono sistemi difensivi che possono funzionare in maniera autonoma e sono già utilizzati in questa modalità operativa quando il tempo utile per reagire efficacemente a un attacco è ritenuto incompatibile con i ritardi insiti nel ciclo percezione-decisione-azione di un operatore umano. Armi autonome offensive già esistenti sono le cosiddette munizioni *loitering*: velivoli senza equipaggio e armati di una testata esplosiva, i quali possono vagabondare per un certo periodo di tempo alla ricerca di un obiettivo da attaccare all'interno di un perimetro prefissato. Insieme all'area geografica concessa per vagabondare, l'operatore umano stabilisce prima della loro attivazione quali siano le categorie di obiettivi che l'arma è abilitata ad attaccare e la durata massima del vagabondaggio (Boulanin e Verbruggen, 2017, p. VII). Il missile israeliano *Harpy* è di tipo *loitering*. Esso può vagabondare per ore in cerca di segnali emessi dalle postazioni radar del nemico, piombando su di esse per distruggerle al momento dell'impatto (Rossi, 2019, p. 8). Le

[96]https://www.army-technology.com/projects/irondomeairdefencemi/.
[97]https://www.army-technology.com/projects/mantis/.

situazioni nelle quali opera *Harpy* possono evolvere rapidamente e in maniera imprevista nella fase di pianificazione dell'azione. Mentre *Harpy* vagabonda sull'area stabilita, un veicolo adibito al trasporto di civili potrebbe avvicinarsi a una postazione radar nemica, oppure la postazione mobile potrebbe spostarsi verso un ospedale. In entrambi i casi, un attacco di *Harpy* alla postazione radar potrebbe avere conseguenze incompatibili con i principi di distinzione e proporzionalità codificati nel diritto internazionale umanitario. E i danni provocati sarebbero difficilmente riconducibili alle responsabilità degli operatori umani, poiché il contesto operativo è cambiato in modo imprevisto dopo l'attivazione del sistema. Pertanto, al fine di esercitare il CUS su un tale sistema, è necessario che un operatore umano possa monitorare costantemente se vi siano oggetti o persone da proteggere nel campo di azione dell'arma e abbia il potere effettivo di autorizzarne l'attacco o di impedirne l'esecuzione.

In definitiva, come mostra la discussione del CUS su sistemi d'arma tanto diversi tra loro per destinazione d'uso, si sta facendo strada nel dibattito accademico, diplomatico e politico, l'idea che si debbano considerare livelli diversi di controllo umano, in funzione della tipologia di arma autonoma e dei suoi contesti operativi (Amoroso e Tamburrini, 2019; IPRAW, 2019).

Si pone allora il problema di distinguere livelli diversi di controllo condiviso tra uomo e macchina nello svolgimento dei compiti critici di selezione e ingaggio di un obiettivo e di assegnare ciascuna arma a un livello che sia appropriato in base alle condizioni etiche e giuridiche esaminate più sopra. La seguente lista di livelli di controllo condiviso (Sharkey, 2016; Amoroso e Tamburrini, 2019) dell'azione tra uomo e macchina (dal livello L1 a L5) – ordinati in base ai privilegi di controllo che a mano a mano si spostano dall'essere umano (che ha il controllo totale a L1 sulle funzioni critiche) alla macchina (che ha il controllo totale a L5) – ci permette di inquadrare meglio il problema:

L1: La selezione dell'obiettivo da attaccare è integralmente effettuata dall'operatore umano;

L2: La selezione dell'obiettivo da attaccare è effettuata dall'operatore umano in base a un ventaglio di opzioni suggerite dal sistema d'arma;

L3: L'operatore umano si limita ad approvare o a rifiutare la scelta del sistema d'arma in merito all'obiettivo da attaccare;

L4: L'operatore umano supervisiona la selezione dell'obiettivo effettuata dal sistema, mantenendo la possibilità di riprendere il controllo ed annullare l'attacco;

L5: L'operatore umano si limita ad attivare il sistema d'arma, definendone la missione nella fase preliminare di pianificazione, senza avere la possibilità di intervenire nella fase operativa.

Abbiamo visto che L5 è un livello troppo sbilanciato a favore della macchina, almeno per quanto riguarda la ricerca dinamica di obiettivi da parte di *Harpy* e di altre munizioni di tipo *loitering*. D'altra parte, i livelli L1-L2, se proposti come i più appropriati per l'esercizio del CUS, sono troppo sbilanciati a favore dell'uomo, almeno

per quanto riguarda l'utilizzazione di sistemi difensivi come *Iron Dome* e *MANTIS*.

Una soluzione differenziata al problema del CUS modulerà opportunamente il controllo umano, lungo i livelli L1-L5, in base alle caratteristiche dei sistemi d'arma e dei loro contesti d'uso. Da un lato, per garantire che gli operatori umani esercitino sempre il CUS, non è sempre necessario escludere l'autonomia dei sistemi d'arma nelle funzioni critiche di selezione e attacco di un obiettivo, come pretenderebbe la soluzione uniforme ad autonomia zero. Da un altro lato, per mantenere il CUS non è sempre sufficiente circoscrivere l'esercizio del controllo umano alla fase di pianificazione dell'azione militare.

Tra le soluzioni differenziate che si possono escogitare e sottoporre al dibattito diplomatico e politico, ci limitiamo a considerare una soluzione che è *differenziata* sui livelli L1-L5, ma *prudenziale* per quanto riguarda le condizioni per il mantenimento del CUS (Amoroso e Tamburrini, 2019).

Vediamo concretamente come potrebbe svilupparsi una tale soluzione differenziata e prudenziale al problema del CUS. Consideriamo innanzitutto la sentinella robotica SGR-A1 (Rossi, 2019), ideata in Corea del Sud per selezionare ed eventualmente aprire il fuoco autonomamente sugli intrusi nella zona demilitarizzata tra le due Coree. Un CUS adeguato su SGR-A1 potrebbe essere esercitato a livello L2, che prevede l'approvazione o il rifiuto di un operatore umano delle proposte del sistema in merito agli obiettivi da attaccare. Un controllo a livello L2 potrebbe essere sufficiente anche per il cosiddetto *deliberate targeting*, purché l'operatore umano abbia informazioni adeguate per decidere se vi siano stati tali cambiamenti nello scenario operativo da richiedere una nuova pianificazione o l'interruzione della missione. Nel caso di *Iron Dome* e di altri sistemi difensivi simili, il CUS potrebbe essere adeguatamente mantenuto al livello L3, che prevede la supervisione di un operatore umano con diritto di veto nelle fasi operative, anche se l'esercizio effettivo del diritto di veto sarà, per forza di cose, fortemente limitato dai tempi rapidi di risposta del sistema.

Più in generale, una soluzione differenziata ma prudenziale, potrebbe articolarsi su due punti principali. Per prima cosa, in assenza di informazioni specifiche al contrario, si possono imporre per default i livelli L1 e al più L2 di controllo umano (*aspetto prudenziale della soluzione*). Potrebbero però essere ammesse eccezioni opportunamente argomentate alla regola default, consentendo di andare oltre L2 nel caso di specifici sistemi d'arma e modalità d'uso, a patto che il CUS non sia minacciato spostando in questo modo i privilegi di controllo verso la macchina (*aspetto differenziato della soluzione*).

Con quali motivazioni si può sostenere la scelta di dare priorità a una regola default che indica il livello L2 di controllo condiviso come quello da adottare in assenza di informazioni al contrario? Le principali motivazioni per questa scelta derivano dalla capacità limitata di prevedere e controllare il comportamento dei sistemi d'arma basati sulle tecnologie dell'IA e della robotica. Questi limiti dipendono dalle capacità stesse

della macchina e dalle sue interazioni con l'ambiente operativo:

- la capacità di apprendere, fondamentale per sviluppare un sistema avanzato dell'IA e della robotica;
- la capacità di vagare (*loitering*) per un periodo esteso di tempo alla ricerca di un obiettivo da attaccare su un territorio che è soggetto a cambiamenti anche repentini delle condizioni osservate al momento di pianificare l'azione dell'arma autonoma;
- la capacità di coordinarsi con altri sistemi in assenza di un controllo centralizzato che l'operatore umano può supervisionare (intelligenza da sciame);
- le interazioni e le conseguenti perturbazioni del comportamento atteso che derivano dai tentativi di jamming, hackeraggio, attacchi cibernetici effettuati con virus o altro tipo di malware informatico;
- le interazioni competitive o cooperative con altri sistemi artificiali, su un campo di battaglia destinato a divenire sempre più informatizzato e robotizzato.

In definitiva, in assenza di prove convincenti che le limitazioni previsionali e di controllo dovute a questi vari fattori non si applicano nel caso di specifiche armi autonome, l'imposizione della regola default offre una protezione prudenziale dalle violazioni degli obblighi deontologici che danno forma ai contenuti del CUS.

In conclusione, ricordiamo ancora che Austria, Cile e Brasile (Austria, 2018) hanno richiesto che al GGE del CCW sia dato il mandato di negoziare uno strumento giuridico vincolante per tutti gli Stati, finalizzato ad "assicurare un controllo umano significativo sulle funzioni critiche dei sistemi d'arma autonomi con effetti letali". Tale proposta ha incontrato le resistenze di alcune potenze militari, ma il vasto consenso internazionale che si è raccolto intorno all'idea del CUS depone a favore della possibilità di pervenire a un trattato internazionale che imponga il CUS su tutti i sistemi di arma. Come punto di partenza di tale trattato, si dovranno evidenziare i problemi etici e giuridici sollevati dall'autonomia dei sistemi d'arma. Queste criticità dovranno diventare parte integrante del Preambolo, fornendo così lo sfondo interpretativo del Trattato. Ma le sfide più importanti per la comunità politica e diplomatica internazionale saranno poste da quella parte del trattato che dovrà affrontare il problema del controllo nelle fasi di pianificazione e di uso dei sistemi d'arma. Se i contenuti di questa parte saranno ispirati a una politica etica e giuridica, che sia differenziata e prudenziale, il trattato dovrà prevedere l'applicazione per default dei livelli più elevati di controllo umano (L1-L2), in tutte quelle circostanze nelle quali non vi sono informazioni tali da giustificare l'introduzione di specifiche eccezioni, da sottoporre all'esame e all'approvazione della comunità internazionale degli Stati nel consesso del CCW o in altre sedi internazionali opportune

Bibliografia

Amoroso D., Sauer F., Sharkey N., Suchman L., Tamburrini G. (2018). *Autonomy in Weapon Systems. The Military Application of Artificial Intelligence as a Litmus Test for Germany's New Foreign and Security Policy.*Berlino: Heinrich Böll Stiftung.

Amoroso D., Tamburrini G. (2017). "The Ethical and Legal Case Against Autonomy in Weapons Systems". In *Global Jurist*, 17(3): 1-20.

Amoroso D., Tamburrini G. (2019). *What makes human control over weapon systems "meaningful"?.* Working paper #4, International Committee for Robot Arms Control (https://www.icrac.net/research/).

Article 36 (2013). *Killer Robots: UK Government Policy on Fully Autonomous Weapons.*Londra, 19 aprile 2013 (http://www.article36.org/weapons-review/killer-robots-uk-government-policy-on-fully-autonomous-weapons-2/).

Austria, Brasile, Chile (2018).*Proposal for a Mandate to Negotiate a Legally Binding Instrument that addresses the Legal, Humanitarian and Ethical Concerns posed by Emerging Technologies in the Area of Lethal Autonomous Weapons Systems (LAWS), submitted to the Group of Governmental Experts on lethal autonomous weapons of the CCW*. Ginevra: ONU (https://undocs.org/CCW/GGE.2/2018/WP.7).

Boulanin V., Verbruggen M. (2017). *Mapping the development of autonomy in weapons systems.* Stoccoloma: SIPRI - Stockholm International Peace Research Institute (https://www.sipri.org/publications/2017/other-publications/ mapping-development-autonomy-weapon-systems).

Crootof R. (2016). "A Meaningful Floor for 'Meaningful Human Control'*".* In *Temple Journal of International & Comparative Law*, 30: 53-62.

DoD (2012). *Autonomy in Weapons Systems*. Directive 3000.09.Washington DC:US Department of Defense: 13–14 (www.dtic.mil/whs/directives/corres/pdf/300009p.pdf).

HRW (2015).Mind the Gap: The Lack of Accountability for Killer Robots. Rapporto pubblicato il 9 aprile 2015. New York: Human Rights Watch(https://www.hrw.org/report/2015/ 04/09/mind-gap/lack-accountability-killer-robots).

ICRC (2014). *Autonomous weapon systems: Technical, military, legal and humanitarian aspects. CCW Expert meeting.* Ginevra: International Committee of the Red Cross (https://www.icrc.org/en/document/report-icrc-meeting-autonomous-weapon-systems-26-28-march-2014).

IPRAW (2019). *Focus on human control*. Report n. 5, Berlin, Agosto 2019 (https://www.ipraw.org/human-control/).

Rossi J.C.(2019). "Un'opera dell'uomo: le macchine autonome letali". In *IRIAD Review – Studi sulla pace e sui conflitti*, n. 5: 2-22.

Sharkey N. (2016). "Staying in the loop: human supervisory control of weapons". In Bhuta N. *et al*. (eds) (2016): 23-38.

UNIDIR (2016). *Statement of the UN Institute for Disarmament Research at the CCW Informal Meeting of Experts on Lethal Autonomous Weapon Systems*, 12 Aprile, (http://www.unidir.org/files/medias/pdfs/unidir-s-statement-to-the-ccw-informal-meeting-of-experts-on-lethal-autonomous-weapon-systems-eng-0-648.pdf.)

Cap. 8
Il ruolo della società civile

8.1 Le iniziative della società civile

Ad oggi le armi completamente autonome non esistono, ma vengono utilizzati molti sistemi di armi con un alto grado di autonomia rispetto ad alcune loro funzioni operative, logistiche e di natura bellica. Per questo motivo già nel 2012 Human Rights Watch (HRW)[98] pubblicò un Rapporto intitolato *Loosing Humanity; the Case Against Killer Robots* in collaborazione con l'*International Human Rights Clinic* (IHRC) dell'Università di Harvard. Il Rapporto analizzava gli aspetti etici di un futuro uso di armi completamente autonome, sottolineando l'assoluta inadeguatezza di tali armi a conformarsi ai principi e alle leggi del diritto Internazionale umanitario che prevedono la protezione dei civili coinvolti in un conflitto. Secondo il Dott. Ron Arkin, professore di robotica presso la School of Interactive Computing, College of Computing del Georgia Institute of Technology, progettare robot capaci di distinguere tra un target miliare e civile (legittimo ed illegittimo) richiederebbe incredibili sfide tecnologiche, considerando che sarebbe già sufficientemente complicato per un robot distinguere un essere umano dall'ambiente circostante, figuriamoci un militare in divisa da un civile, e cosa più importante un combattente da un non-combattente. L'introduzione di armi completamente autonome in un conflitto, secondo il Rapporto, condurrebbero a difficili interpretazioni delle misure di precauzione e della regola della proporzionalità stabiliti giuridicamente dalla dottrina dello *jus in bello* che stabilisce non solo che i combattenti devono, nella condotta delle operazioni militari, risparmiare la vita dei civili, ma che addirittura debbano prendere misure adeguate per evitare vittime tra i non combattenti.

Alla luce dei dubbi espressi sia a livello accademico, sia a livello di organizzazioni umanitarie internazionali come HRW, nell'aprile 2013 venne lanciata la *Campaign To Stop Killer Robots*, (la Campagna per fermare i *killer robots*) costituita, ad oggi, da una coalizione di circa centoquaranta Organizzazioni Non Governative che lavorano congiuntamente affinché vengano messe al bando le armi totalmente autonome. Ciò che la Campagna chiede ai governi di tutto il mondo ed alle Nazioni Unite è quello ratificare un Trattato internazionale che vieti l'utilizzo di questi sistemi d'arma e che venga tassativamente mantenuto il controllo umano nel ciclo decisionale (targeting ed attacco).

A sostegno della Campagna circa 4.500 scienziati (tra cui Stephen Hawking, Yoshua Bengio[99] e molti esperti di intelligenza artificiale) nel 2015 hanno aderito alla

[98]http://www.law.harvard.edu/programs/hrp/ihrc.html (13 marzo 2020).
[99] Bengio nel 2018 ha ricevuto, insieme a Geoffrey Hinton e Yann LeCun, il Turing Award dell'Association

sottoscrizione di una lettera[100] contro i sistemi d'arma letali autonomi. Nel testo viene innanzitutto posta l'attenzione sul fatto che la tecnologia dell'intelligenza artificiale (IA) ha raggiunto un punto in cui lo spiegamento di tali sistemi è – praticamente se non legalmente – fattibile in anni, non decenni, e la posta in gioco è molto alta: "Se una grande potenza militare avanza con lo sviluppo delle armi AI, una corsa globale agli armamenti è praticamente inevitabile e il punto finale di questa traiettoria tecnologica è ovvio: le armi autonome diventeranno i Kalashnikov di domani".

Nella lettera viene, però, messa in evidenza anche come la tecnologia abbia molti aspetti positivi, ma che, allo stesso tempo, l'uso dell'IA, applicata in campo militare, costituirebbe una grave minaccia e che si rende necessario un divieto di mettere in campo bellico tali armi senza un significativo controllo umano.

La società civile, nonché molti ricercatori e scienziati di tutto il mondo hanno evidenziato una serie di criticità etiche, legali, operative, di proliferazione e di sicurezza a livello globale legate all'eventuale utilizzo di tali sistemi di armi in guerra:

1. Una macchina non dovrebbe mai prendere decisioni totalmente autonome circa la vita e la morte di un essere umano.
2. Le armi autonome non sarebbero in grado di rispettare i principi di proporzionalità, distinzione e necessità militare, che sono alla base del diritto umanitario internazionale e che solo la coscienza umana è in grado di soddisfare.
3. Tali armi sono relativamente economiche e facili da copiare. Esiste un rischio molto elevato che, una volta prodotte, potrebbero facilmente essere utilizzate anche da attori non-statali. Nel 2016 lo Stato Islamico ha portato a termine il primo attacco realizzato con un drone nel nord dell'Iraq, uccidendo due combattenti Peshmerga (Warrick, 2017). Nel 2017 il gruppo terroristico ha annunciato la formazione del primo "Unmamned Airfcraft of the Mujahedden."
4. Molti governi speculano poi sulla convinzione che tali armi porterebbero ad un numero sempre minore di vittime, ma la percezione di un rischio minore potrebbe portare a preferire una soluzione militare rispetto ad una diplomatica, stravolgendo quindi il concetto di Guerra così come teorizzato da SunTzu per il quale "il governante accorto è solo colui che risulta vittorioso non combattendo nessuna Guerra".
5. Le armi totalmente autonome rischiano di creare un pericoloso gap riguardo alle responsabilità nel caso, ad esempio, dell'uccisione di un civile. Il MIT (Massachusetts Institute of Technology) ha dato il via, nel 2016, ad un progetto di ricerca chiamato *Moral Machine* che ha coinvolto circa 2.3 milioni di persone in 233 paesi. Il Progetto si basa sulla somministrazione di un *test* nel quale vengono proposti "dilemmi

for Computing Machinery, il riconoscimento scientifico più alto per gli informatici (il Premio Nobel dell'Informatica).

[100]https://futureoflife.org/open-letter-autonomous-weapons/?cn-reloaded=1.

morali" su situazioni di possibili incidenti nei quali una delle persone coinvolte deve essere "sacrificata". Al di là delle perplessità sollevate da alcuni esperti sul *test* in questione, ciò che emerge è che i principi morali cambiano da paese a paese così come le preferenze soggettive risultano profondamente condizionate dalle variabili sociali, culturali, economiche e religiose; in Giappone ed in Finlandia le persone sono più propense a sacrificare un pedone che non attraversa sulle strisce, mentre lo stesso pedone avrebbe molte più possibilità di salvare la propria vita se vivesse in Nigeria oppure in Pakistan (Spagnolo, 2019). La situazione si complicherebbe ulteriormente in un ipotetico scenario di guerra, dove, per esempio, un sistema d'arma autonoma in avaria, dal valore di milioni di dollari, dovrebbe scegliere se schiantarsi su una montagna oppure cercare di atterrare in un centro di una cittadina con il rischio di uccidere molti civili.

6. L'impiego sempre più preponderante della tecnologia nei conflitti armati rischia di portare ad una corsa agli armamenti con effetti profondi sulla pace e sulla sicurezza mondiali, permettendo, inoltre, ai dipartimenti della difesa ed alle aziende produttrici di questi sistemi d'arma di avere libero accesso a milioni di dati personali con un potere di controllo sulla vita di milioni di persone e con il rischio di subire attacchi *cyber* di natura criminale in grado di generare una catastrofica guerra planetaria. Nel giugno del 1997 25 funzionari della NSA (National Security Agency) "bucarono" le reti informatiche del Dipartimento della Difesa statunitense. Il Read Team (nome in codice dei venticinque funzionari) penetrò il sistema di sicurezza americana in soli quattro giorni (Accoto, 2019).
7. L'uso di tali armi in un conflitto non può escludere la possibilità di un errore accidentale che porterebbe ad una rapida escalation con conseguenza imprevedibili e pericolose (Pax, 2019).

Nel 2013, grazie alla forte pressione della società civile, si è svolto un primo incontro informale presso le Nazioni Unite di Ginevra, mentre nel 2016 molti stati che facevano già parte della CCW (*Convention on Certain Conventional Weapons*), entrata in vigore nel 1983 con l'obiettivo di limitare l'utilizzo di certe armi durante i conflitti armati, hanno deciso di fare includere in agenda anche le armi totalmente autonome.

Nel 2017 oltre duecento aziende tecnologiche, tra cui Clearpath Robotics e Google, si sono impegnate a non supportare le ricerche dell'IA in ambito militare chiedendo inoltre alle Nazioni Unite di vietare la produzione di tali armi. Nel giugno 2018 quattromila dipendenti di Google hanno aderito ad una petizione che chiedeva all'azienda di interrompere un controverso programma di ricerca, il Progetto Maven, in collaborazione con il Pentagono, che implicava l'uso dell'Intelligenza Artificiale in campo militare. La lettera, indirizzata direttamente all'Amministratore Delegato Sundar Pichai, chiedeva esplicitamente di cancellare tale progetto e di provvedere a formulare una *policy* che impegnasse Google a non collaborare più allo sviluppo di tecnologie legate al mondo della difesa (Campaign To Stop Killer Robots, 2018). Anche Antonio

Guterres, Segretario Generale delle Nazioni Unite, in un discorso tenuto nella sede delle Nazioni Unite di Parigi nel 2019, si è espresso a favore della realizzazione di un Trattato internazionale che vieti l'uso dei killer robot dichiarando che "sono politicamente inaccettabili e moralmente ripugnanti quelle macchine che hanno potere di decidere sulle vite degli esseri umani e dovrebbero essere vietate da una legge internazionale" (Pax for Peace).

Il 2019 ha segnato diversi e importanti traguardi nell'ambito dell'impegno globale della società civile su questo tema poiché ben settantanove iniziative (cinquantotto progetti nazionali, sedici regionali e cinque incontri internazionali) di sensibilizzazione sono state avviate attraverso il programma di sovvenzione della Campagna che hanno coinvolto governi a livello nazionale ed opinione pubblica sulla necessità, sempre più urgente, di avere risposte sulle sfide etiche, legali, tecniche e sociali sollevate dal continuo sviluppo di tali armi. Tali attività si sono svolte in trentaquattro paesi del mondo (tra cui Argentina, Francia, Italia, Camerun, Pakistan etc.). Le iniziative regionali sono state organizzate dai membri della Campagna in tutto il mondo e numerose attività sono state svolte presso enti regionali come l'Unione Africana, l'Organizzazione di Stati Americani e la comunità di sviluppo dell'Africa Australe. Gli incontri internazionali sono stati ospitati a Tokyo, Anversa, Nur-Sultan, Medellin ed Islamabad etc. (Stop Killer Robots, 2019) con la partecipazione sia di esperti, sia del mondo dei media, sia dell'opinione pubblica.

Infine, a febbraio 2020 si è svolto, a Buenos Aires, il meeting internazionale della Campagna, dove sono state suggerite nuove strategie per stimolare maggiormente il dibattito internazionale ed informare sempre più dettagliatamente l'opinione pubblica. Pax ha pubblicato il Rapporto *Save your university from killer robots* (Pax, 2020*)*, in cui viene evidenziato il fondamentale ruolo delle università nella promozione di tali tematiche, creando discussioni all'interno delle classi sull'impatto delle nuove tecnologie nella nostra società, nonché l'invito ad organizzare, all'interno delle università, eventi sul tema relativo ai killer robot.

8.2 Conclusioni

L'impegno della società civile, nell'ambito di un tema così complesso, ha l'obiettivo innanzitutto di sollecitare i governi mondiali e le istituzioni ed organizzazioni intergovernative a votare un mandato che si possa tradurre, il più velocemente possibile, in un Trattato con valore legale. Inoltre la creazione di una rete internazionale formata da esperti di discipline sia scientifiche sia umanitarie ha permesso di pubblicare e fare circolare documenti e report tradotti in molte lingue consultabili sia da esperti, ma soprattutto da comuni cittadini che hanno la possibilità di venire a conoscenza di queste tematiche e delle sue criticità rendendo partecipi attivamente

migliaia di persone in ogni paese del mondo sui temi della pace e del disarmo umanitario.

Bibliografia

Accoto C. (2019). *Mondo ex Machina. Cinque brevi lezioni di filosofia dell'automazione.* Milano*:* Egea Editore.

Human Rights Watch (2012). *Loosing Humanity. The Case Against killer Robots*. Harvard University.

Kurzweil R. (1990). *The Age of Intelligent machines*. Cambridge: MIT Press.

McCarthy J., Minski M.L., Rochester N., Shannon C.E. (1995).*A proposal for the Dartmouth summer research project on artificial intelligence*(http://jmc.stanford.edu/articles/ dartmouth/dartmouth.pdf).

Melegari A. (2018). "Come dovrebbero ragionare le armi del futuro?". In *Analisi e Difesa,* 30 ottobre 2018.

PAX (2019). *State of AI,* n. 689, April.

Pax for Peace NL (2019). *Killer Robots. What are they and what are the concerns*, 689, August.

Rossi J.C. (2019). "Intelligenza Artificiale e robotica alla guerra". In *IRIAD Review. Studi sulla pace e sui conflitti*, 5, 2019: 2-22.

Spagnolo A. (2018). "La regolamentazione delle armi autonome: letteratura o diritto?". In *Torino World Affairs Institute*.

Surber R. (2018). *Artificial Intelligence: Autonomous Technology (AT), Lethal Autonomous Weapons (LAWS) and Peace Time Threats*. ICT For Peace Foundation and the Zurich Hub for Peace Foundation.

Warrick J. (2017). "Use of weaponized drones by ISIS spurs terrorism fears". In *The Washington Post*, 21 febbraio 2017.

Sitografia

http://moralmachine.mit.edu/
https://www.paxforpeace.nl/our-work/programmes/killer-robots
http://www.law.harvard.edu/programs/hrp/ihrc.html
https://futureoflife.org/open-letter-autonomous-weapons/?cn-reloaded=1
https://www.stopkillerrobots.org/2018/05/google/
https://www.emergency.it/cosa-facciamo/afghanistan/
http://www.intelligenzaartificiale.it/machine-learning/
https://www.iai.it/sites/default/files/pi_n_0081.pdf
https://www.paxforpeace.nl/our-work/programmes/killer-robots
https://www.europarl.europa.eu/doceo/document/TA-8-2018-0341_IT.html?redirect
http://moralmachine.mit. edu/)

Cap. 9
Considerazioni finali. Linee d'azione e di controllo a livello nazionale e europeo: possibilità e potenzialità

9.1 La Rome Call for AI Ethics

"L'intelligenza artificiale offre un potenziale enorme quando si tratta di migliorare la convivenza sociale e benessere personale, aumentando le capacità umane e abilitando o facilitando molti compiti che possono essere effettuati in modo più efficiente ed efficace. Tuttavia, questi risultati non sono affatto garantiti".

Questo è il preambolo del manifesto *Rome Call for AI Ethics*, firmato a Roma nel febbraio 2020 dal presidente della Pontificia Accademia per la Vita, monsignor Vincenzo Paglia, dal vicepresidente dell'Ibm John Kelly III, dal presidente di Microsoft Brad Smith, dal direttore generale della Fao Dongyu Qu e dalla ministra dell'Innovazione tecnologica e la digitalizzazione, Paola Pisano. In tale documento si evidenziano l'importanza e le criticità dell'IA, che certamente sta trasformando, se non rivoluzionando sempre più la nostra società offrendo molteplici e rilevanti opportunità. Per comprenderne la portata basta solo pensare alle grandi applicazioni nel settore sanitario nell'ambito della terapia intensiva, nel supportare i medici nell'elaborazione delle diagnosi, nella previsione della diffusione delle malattie, nel personalizzare al massimo le cure, nelle diagnosi a distanza e così via.

9.2 L'IA nel settore degli armamenti

L'IA, però, è in via di larga e crescente applicazione non solo nel settore civile, ma anche in quello militare. In quest'ultimo campo, le applicazioni dell'IA ai sistemi d'arma, seppur relativamente agli esordi, sono di fatto già una realtà in ambito industriale e militare presso numerosi stati, dagli USA alla Russia, dalla Francia alla Gran Bretagna, dalla Corea del Sud ad Israele.

Svariati sistemi d'arma (come munizioni *loitering*, sistemi d'arma ravvicinata o a corto raggio (CIWS), aerei da combattimento senza equipaggio UAV, munizioni guidate di precisione, veicoli terrestri senza equipaggio - UGV e veicoli marini senza equipaggio - UMV) sono sottoposti ad applicazioni sempre più avanzate tese a realizzare macchine belliche capaci non solo di operare in una prospettiva di crescente autonomia, ma anche in uno spazio temporale assai più breve di quello umano. Se molti di questi sistemi d'arma sono ancora in via di R&S o sotto forma di prototipi, altri sono già operativi soprattutto nel settore difensivo, come l'S-400 russo (dal 2007) o il MIM 404 Patriot statunitense.

Ingenti risorse già da alcuni anni vengono riversate dai governi interessati in questo

settore, che appare destinato ad una crescita esponenziale, come è stato evidenziato nel cap. 3, coinvolgendo di conseguenza sempre di più anche le industrie e il mondo della ricerca. Nel complesso le previsioni per l'impatto economico dell'IA (civile e militare) si aggirano intorno ai $ 13 trilioni entro il 2030, secondo un rapporto del settembre 2018 del McKinsey Global Institute (Wladawsky-Berger, 2018).

9.3 Criticità e rischi dell'IA nel settore bellico

Se la tecnologia basata sull'IA è, come già detto, importantissima in tanti settori, in quello della difesa essa pone però alcune criticità (Tonin, 2019). L'IA presenta, almeno allo stato attuale, ancora forti limiti. I casi accennati in precedenza nel cap. 2, come quelli che hanno utilizzato la metodologia di *adversarial testing,* hanno messo in rilievo come la rete neurale possa essere indotta in errore, scambiando uno scuolabus e un cane per un'ostrica o una tartaruga per un fucile.

Ciò evidenzia anche la parziale affidabilità dell'IA ad oggi e questo nel settore delicato come quello della difesa è un problema non secondario (come è stato documentato nel cap. 2). Analogamente l'analisi del processo di *machine learning* ha fatto emergere alcuni punti critici che non possono essere trascurati, tanto che risulta assai rischioso affidarsi a tali tecniche come *unico strumento* decisionale, stando almeno ai livelli di ricerca attuali.

Sono, tra l'altro, emerse vulnerabilità dei sistemi di IA e ML connesse sia a *exploratory attacks*, sia a *causative attacks*, sia a *catastrophic forgetting*. Emergono elevati rischi connessi sia ad un'azione offensiva avversaria che con varie modalità può indurre in errore il sistema, sia a malfunzionamenti interni e difficilmente prevedibili dell'IA dovuti ai propri processi di autoapprendimento e al complesso funzionamento degli algoritmi (cap. 4).

L'interesse destato dai LAWS presso vari governi e stati maggiori è connesso a molteplici motivi. In primo luogo, essi potrebbero offrire una maggiore precisione negli attacchi rispetto a quella umana, nonché un'elevata rapidità nella reazione ad un attacco subito. In secondo luogo, essendo delle macchine non soggette allo stress emotivo a cui è esposto l'uomo in una situazione conflittuale, potrebbero ridurre i danni collaterali rispettando meglio il diritto umanitario internazionale. In terzo luogo, tali macchine potrebbero essere programmate per essere utilizzate solo se esposte ad un evidente e certo atto aggressivo, decisione d'ingaggio che non potrebbe essere fatta per la vita del soldato in quanto lo esporrebbe in misura maggiore al pericolo.

D'altro canto, si può considerare che le possibili minori perdite di vite dei propri militari potrebbero incentivare sia la disponibilità dell'opinione pubblica, sia quella dei decisori politici verso nuovi conflitti armati, fondandosi sulla convinzione di una guerra

combattuta solo dalla macchine[101]. Per di più, tale disponibilità potrebbe avere delle conseguenze inattese per un conflitto ipotizzato a perdite zero che potrebbe evolvere in modo del tutto imprevisto. Anche se la locuzione ciceroniana "historia magistra vitae" presente nel *De Oratore* non appare trovare conferma nelle azioni della maggior parte dell'umanità, basta ricordare che anche il primo conflitto mondiale si avviò nella convinzione di una guerra breve, in quanto tutti i contendenti erano convinti della superiorità tecnologica dei propri arsenali.

Tra l'altro, le maggiori potenze industriali (e militari) potrebbero essere indotte a ritenersi facili vincitrici di guerre contro Paesi non dotati di LAWS, ma la supremazia tecnologica comunque non ha sempre garantito la vittoria, come hanno evidenziato i conflitti del Vietnam e dell'Afghanistan.

Inoltre, dati i costi ridotti della robotica e dell'IA, potrebbero svilupparsi non solo una nuova corsa agli armamenti coinvolgente un numero elevato di stati, ma anche una diffusione di tali sistemi d'arma presso forze armate irregolari e formazioni terroristiche, con conseguenze imprevedibili sulla sicurezza interna ed esterna. Se poi consideriamo anche i progressi nel settore delle stampanti 3D e i pericoli connessi all'evoluzione tecnologica e dei materiali, i rischi appaiono ancora più diffusi.

Infine, non può essere trascurata la questione della progressiva emarginazione del ruolo attivo dei militari nell'ambito di un conflitto basato sui LAWS, che andrebbero gradualmente a sostituirli sia in situazioni non conflittuali sia durante scontri armati, lasciando il posto a tecnici informatici, dotati di una preparazione e di una formazione assai differenti. Ovviamente non si pone la questione in termini di luddismo, ma sul ruolo fondamentale che svolgono i militari, con capacità di ampio discrimine, nel momento in cui si trovano ad agire sul teatro operativo. Anche nel caso di regimi autoritari o dittatoriali una diffusione di tali sistemi d'arma potrebbe consentire ai governanti azioni ed interventi indiscriminati che potrebbero, teoricamente e in determinate situazioni, essere invece rifiutati dai membri delle forze armate e di sicurezza.

[101] Non a caso il presidente Putin nel settembre 2017 sottolineò il rischio che "chi diventerà leader in questa sfera sarà il sovrano del mondo", affermando poi successivamente, nel novembre 2019 *all'evento Artificial Intelligence Journey* a Mosca, l'intenzione di far diventare la Russia leader mondiale nel settore, compreso quello della difesa. Vedi https://tg24.sky.it/mondo/2017/09/01/putin-intelligenza-artificiale.html; Rovelli M. (2017), "Putin sull'intelligenza artificiale: «Chi sviluppa la migliore, governa il mondo». Musk rilancia l'allarme", https://www.corriere.it/tecnologia/economia-digitale/17_settembre_04/putin-sull-intelligenza-artificiale-chi-sviluppa-migliore-governa-mondo-musk-rilancia-l-allarme-c2a46c9c-916f-11e7-8332-148b1c29464d.shtml.

9.4 Gli aspetti giuridici

Certamente l'uso di sistemi d'arma autonomi comporterebbe una trasformazione radicale nella gestione dei conflitti, con ineludibile sovvertimento del quadro in ambito etico, giuridico, comportamentale e operativo.

D'altronde, come è stato rilevato nel cap. 5, la società civile ha avvertito i rischi e i pericoli di una diffusione incontrollata di tale tecnologia ai sistemi d'arma, come è emerso sia dall'indagine IPSOS del 2018 su 26 Paesi, sia da quella dell'Archivio Disarmo del febbraio 2019. In esse gli italiani hanno evidenziato una contrarietà rilevante (58% nel primo caso, 68% nel secondo). Né vanno trascurate le numerose prese di posizione della comunità scientifica già dal 2012 sino ad arrivare alla lettera aperta del 2015 firmata da migliaia di ricercatori dell'IA e della robotica e ad altri successivi appelli da imprenditori dei settori dell'informatica, della robotica e dell'intelligenza artificiale, da gruppi di loro dipendenti, da ricercatori operanti in settori affini (cap. 7).

Dal punto di vista giuridico, i LAWS devono comunque fare riferimento sia al rispetto delle norme del diritto internazionale umanitario, sia al mantenimento della catena delle responsabilità nelle azioni belliche, sia alla questione del rispetto della dignità degli esseri umani coinvolti nelle azioni belliche. E come è stato messo in evidenza (cap. 1), non solo è difficile distinguere un nemico fuori combattimento ai sensi di quanto prevede il Protocollo del 1977, art. 41, aggiuntivo alle Convenzioni di Ginevra del 1949 (principio di distinzione), ma anche applicare il principio di proporzionalità (Protocollo I del 1977, art. 51.5.b).

Va considerato anche che il diritto internazionale è progredito nel tempo e ha previsto la repressione dei crimini internazionali attraverso la Corte Penale Internazionale, mettendo sempre più in evidenza la responsabilità della condotta militare, sin dal processo di Norimberga. Questo principio ovviamente non può essere ignorato dall'attivazione di nuove armi, macchine certo non responsabili come lo è l'uomo. La responsabilità di un errore letale del LAWS (l'abbattimento di un aereo civile o la distruzione di uno scuolabus, ad esempio) non potrebbe essere addossata giuridicamente alla macchina. Dovrebbe essere indirizzata al programmatore dell'algoritmo, alla ditta costruttrice, all'azienda dell'indotto subfornitrice, agli addetti all'installazione, al personale incaricato della manutenzione, a chi ne ha deciso l'acquisto e la dislocazione, a chi ha approvato politicamente l'adozione, ad un'azione avversaria di disturbo, ad eventi imprevisti? La catena delle responsabilità diviene talmente vaga ed indeterminata al punto tale da porre più di un interrogativo sull'adozione indiscriminata di tali armi.

9.5 Il dibattito e le proposte a livello internazionale

Pertanto, le sfide poste dai LAWS sul piano della sicurezza internazionale, come già detto, hanno suscitato nella società civile e nella comunità scientifica una vasta preoccupazione, che si è tradotta in decise prese di posizione, spingendo anche i governi a confrontarsi in ambito multilaterale, come già detto nei cap. 7 e 8. Questo sta avvenendo in sede CCW a Ginevra, dove il dibattito tra le varie posizioni (chi assolutamente contrario, chi decisamente favorevole, chi teso a cercare uno strumento operativo concreto) ha portato ad una serie di passi importanti.

Dal lungo dibattito internazionale presso la CCW sembra emergere il concetto del controllo umano significativo (formula "CUS"), che appare trovare un certo consenso presso le delegazioni governative, anche perché evita di disperdersi nel dibattito su cosa sia un LAWS concentrandosi invece sul suo controllo e sui suoi diversi livelli, come prospettato da Sharkey (2016) e da Amoroso e Tamburrini (2019) secondo lo schema già indicato:

L1: La selezione dell'obiettivo da attaccare è integralmente effettuata dall'operatore umano;

L2: La selezione dell'obiettivo da attaccare è effettuata dall'operatore umano in base a un ventaglio di opzioni suggerite dal sistema d'arma;

L3: L'operatore umano si limita ad approvare o a rifiutare la scelta del sistema d'arma in merito all'obiettivo da attaccare;

L4: L'operatore umano supervisiona la selezione dell'obiettivo effettuata dal sistema, mantenendo la possibilità di riprendere il controllo ed annullare l'attacco;

L5: L'operatore umano si limita ad attivare il sistema d'arma, definendone la missione nella fase preliminare di pianificazione, senza avere la possibilità di intervenire nella fase operativa.

In sede di CCW la proposta di alcuni paesi per un mandato al GGE di negoziare uno strumento giuridico vincolante per tutti gli Stati, volto ad "assicurare un controllo umano significativo sulle funzioni critiche dei sistemi d'arma autonomi con effetti letali" appare dunque aver riscosso un vasto consenso, anche se alcune potenze militari sembrano dissentire. Esso potrebbe pertanto portare ad un trattato internazionale teso normativamente a far prevalere i livelli più elevati di controllo umano, come evidenziato nel cap. 7.

9.6 L'azione italiana in sede multilaterale

Come riportato (cap. 6), anche se da parte italiana non è stato prodotto nessun documento in merito, l'azione diplomatica che si è svolta a livello CCW ha puntato verso l'obiettivo di comuni dichiarazioni politiche, seppur non vincolanti, ma in grado comunque di coinvolgere i paesi più restii ad impegnarsi in tale prospettiva.

È significativo in tal senso, già nel 2018, l'intervento del Rappresentante Permanente italiano alla Conferenza sul Disarmo, Amb. Gianfranco Incarnato, con cui ha sostenuto l'applicazione del diritto internazionale umanitario a tutti i sistemi d'arma (compresi i LAWS) in correlazione con l'art. 36 del Protocollo Addizionale alle Convenzioni di Ginevra del 1949. Non va dimenticato che Codice dell'Ordinamento Militare italiano del 2010 prevede che ogni sistema d'arma nella fase di studio, sviluppo, acquisizione o adozione rispetti il diritto internazionale umanitario. Per di più, nel nostro ordinamento le decisioni in merito ai nuovi sistemi d'arma passano anche attraverso il vaglio e la delibera delle Commissioni parlamentari competenti.

Inoltre, il ministro della difesa Guerini ha concordato sulle "perplessità sull'utilizzo spinto dei dispositivi di intelligenza artificiale", aggiungendo poi "la necessità di individuare un adeguato sistema giuridico entro il quale poter collocare la robotica autonoma". Alla luce di queste considerazioni anche di un importante esponente del nostro governo, l'azione italiana a livello internazionale, pur essendo fuori dal gruppo ONU di esperti governativi (cyber GGE), ma dentro l'Open Ended Working Group (OEWG), in base alla formula CUS, che, attraverso la gradualità dei suoi livelli, consentirebbe di poter raccogliere un vasto consenso, può svolgere un ruolo importante.

Come abbiamo visto, vi sono infatti posizioni articolate, alcune tese a formule più vincolanti (ma con il pericolo di esclusioni di alcuni importanti attori), altre volte a creare comunque i più vasti accordi possibili, finalizzati anche a recuperare in qualche modo i più esitanti. Nell'agosto 2019, l'ultimo meeting del Gruppo di Esperti Governativi sui LAWS è riuscito a raggiungere il consenso su un rapporto finale, che comprende la definizione di undici principi guida nel campo dello sviluppo delle LAWS condivisi da tutti gli Stati partecipanti. Successivamente la Conferenza degli Stati parte alla CCW (13-15 novembre 2019), sulla base del report finale, ha deciso di supportare i principi guida proposti dalla GGE.

Il nuovo assetto dell'Unione Europea, con la recente uscita della Gran Bretagna, può comunque offrire uno spazio di azione importante per il nostro paese, in sinergia con Francia e Germania (vedi il Trattato franco-tedesco di Aquisgrana o l'*European Defence Initiative*), proprio alla luce degli impegni per una più attiva politica di difesa comune che l'UE sta portando avanti da circa un ventennio, cercando di sviluppare ulteriormente il progetto europeo di Altiero Spinelli, Jean Monnet, Robert Schumann e Konrad Adenauer.

Infatti, in tal senso una tappa importante è stata la pubblicazione dell'*European Global Strategy* (2016) seguita poi nel biennio successivo dalla *Permanent Structured Cooperation* (PESCO), dall'approvazione dell'*European Defence Action Plan*, con il Fondo Europeo per la Difesa (*European Defence Fund*, con 13 miliardi per il periodo 2021-2027), con il piano di investimento per le piccole e medie imprese (SMEs) e con un complesso di misure per rafforzare il mercato europeo della difesa. Tutte queste

misure sin da ora offrono complessivamente strumenti utili per indirizzare la produzione nel settore della difesa e, di conseguenza, anche eventuali programmi comunitari di R&S nel settore potrebbero essere impostati in campo industriale per essere finalizzati a rispettare la sequenza L1-L5 suddetta e i principi suaccennati.

Né va dimenticata la risoluzione del Parlamento Europeo del 27 febbraio 2014 con cui si è assunta una posizione nettamente contraria alle armi autonome, in cui veniva richiesto di "vietare lo sviluppo, la produzione e l'impiego di armi completamente autonome che consentono di sferrare attacchi senza alcun intervento umano", nonché la Risoluzione del Parlamento Europeo del 12 settembre 2018 del 2018 successiva all'intervento dell'ex alto rappresentante dell'UE Federica Mogherini. Una posizione più prudente è stata quella del Consiglio Ue del novembre 2016 con la quale si richiedeva una regolamentazione ispirata dal diritto internazionale e in particolare dal diritto internazionale umanitario. Comunque sia, appare evidente che in ambito comunitario vi sono spazi importanti per un'azione incisiva sulla questione dell'IA applicata ai sistemi d'arma.

Anche l'Alleanza Atlantica da tempo sta affrontando i problemi connessi all'uso di queste nuove tecnologie sia relativamente all'impatto dell'IA sull'informazione e sul supporto decisionale e i sistemi robotici autonomi, sia in relazione alle implicazioni strategiche, sia in rapporto alle relative sfide tecniche e non (Tonin, 2019), anche con lo scopo di mantenere posizioni di leadership. È interessante notare che il ministro sloveno della difesa Matej Tonin, rapporteur allo Science And Technology Committee (STC) - Sub-Committee on Technology Trends and Security (STCTTS) sottolinea "che tutti gli sforzi di duplice uso e di IA militare dovrebbero, tuttavia, affrontare tutte le questioni etiche, legali e sociali fin dall'inizio, comprese le considerazioni sulla privacy e la definizione di un adeguato coinvolgimento umano nelle decisioni sull'uso della forza", evidenziando l'importanza delle decisioni politiche dato che "le forze armate alleate da sole non saranno in grado di risolvere le sfide specifiche dell'intelligenza artificiale".

Anche nel corso del 34° Think-Tank for Information Decision and Execution (TIDE) Sprint svoltosi dal 21 al 25 ottobre 2019 a Virginia Beach, negli Stati Uniti, si è discusso sia delle implicazioni, delle opportunità e dei rischi per i militari, sia dell'impatto dell'intelligenza artificiale sulle persone, sulla società, e delle relative scelte etiche relative all'uso militare del'IA[102].

Pertanto, pure in ambito NATO si possono fare concreti passi analoghi a quelli effettuabili in sede europea, seppur in presenza di paesi (come gli Stati Uniti, la Gran Bretagna e la Francia) molto impegnati nei programmi connessi ai LAWS. L'azione

[102] "Artificial Intelligence - A Game Changer for the Military", Oct 25, 2019https://www.act.nato.int/articles/artificial-intelligence-game-changer-military.

italiana nelle varie sedi multilaterali, in primis l'ONU, l'UE, la NATO e la CCW di Ginevra (UN, 2019), può effettivamente svolgere, come si è visto, un ruolo significativo.

Dato che non serve un'alta intensità di capitale per ricerche e sviluppo in ambito IA ed è possibile quindi per molti paesi (compresa l'Italia) impegnarsi significativamente nel settore, appare necessaria una conseguente assunzione di responsabilità politica adeguata alle sfide e alle prospettive analizzate. L'IA applicata al settore della difesa può certamente essere un'opportunità, ma presenta ad oggi elevate problematicità che raccomandano un'estrema prudenza nella sua adozione e nelle conseguenti scelte politiche volte a realizzare accordi internazionali adeguati ai rischi evidenziati (Klare, 2019).

L'Italia, in tal senso, può svolgere un ruolo importante di mediazione e di raccordo tra le diverse posizioni dei governi evidenziando le criticità di tale tecnologia e i rischi conseguenti, al fine di giungere ad un accordo internazionale condiviso in grado di evitare una pericolosa deriva anarchica nella proliferazione di questi nuovi sistemi d'arma.

Bibliografia

Battaglia P. (2018).*Intelligenza artificiale: dagli automi ai robot «intelligenti».* Milano: RCS MediaGroup.

Boden M. A. (2019). *L'intelligenza artificiale*. Bologna: Il Mulino.

Brzozowski A. (2019)**.** *No progress in UN talks on regulating lethal autonomous weapons,* 10 December.

Géhin L. (2020). *L'impression 3D est-elle un vecteur de prolifération illicite des armes à feu?* GRIP - Groupe de Recherche et d'Information sur la Paix et la sécurité, Bruxelles, 11 février 2020.

Hénin S. (2019). *AI: intelligenza artificiale tra incubo e sogno.* Milano: Hoepli.

Iaria A. (*2019).* "Superare i confini globali: le armi da fuoco nel dark web". In *IRIAD Review, n. 3, marzo:* 108-128.

Kaplan J**.** (2018). *Intelligenza artificiale: guida al futuro prossimo***.** Roma: Luiss University Press.

Klare M.T. (2019). *Autonomous Weapons Systems and the Laws of War*, Arms Control Association, March (https://www. armscontrol.org/act/2019-03/features/autonomous-weapons-systems-laws-war).

Rome Call for AI Ethics, Roma, 28 febbraio 2020, http://www.academyforlife.va/content/dam/pav/documenti%20pdf/2020/CALL%2028%20febbraio/AI%20Rome%20Call%20x%20firma_DEF_DEF_.pdf.

Tonin M. (2019). *Artificial Intelligence: Implications For Nato's Armed Forces*. NATO Parliamentary Assembly, Science And Technology Committee (STC) Sub-Committee on Technology Trends and Security (STCTTS), 13 October

UN (2019).*Autonomous weapons that kill must be banned, insists UN chief*, (dichiarazione di A. Guterres), 25 March 2019.(https://news.un.org/en/story/2019/03/1035381).

Wladawsky-Berger I. (2018). "The Impact of Artificial Intelligence on the World Economy". In *Wall Street Journal*, 16 November.

GLI AUTORI

Fabrizio Battistelli è professore ordinario di sociologia presso il Dipartimento di scienze Sociali ed Economiche della Sapienza Università di Roma. Autore di numerosi studi sugli aspetti sociali della sicurezza internazionale e interna, è presidente dell'IRIAD, che ha cofondato nel 1982.

Sofia Bertieri, laureatasi in Sviluppo economico e Cooperazione Internazionale a Firenze, successivamente dottoressa magistrale in Scienze Strategiche presso la Scuola di Applicazione di Torino, frequenta il Master di II livello in Diritto Internazionale Umanitario e dei Conflitti Armati presso il Centro Alti Studi per la Difesa. Collabora con l'IRIAD dal 2019.

Francesca Farruggia, Segretario Generale dell'IRIAD, Dottore di ricerca in "Sistemi sociali, organizzazione e analisi delle politiche pubbliche" presso il Dipartimento di Scienze sociali ed economiche dell'Università di Roma "La Sapienza", è professore a contratto di Sociologia della security presso Unitelma-Sapienza.

Barbara Gallo, giornalista pubblicista dal 2016, si è laureata in Sociologia all' Università di Roma "La Sapienza", collabora con l'IRIAD dal 2015 come ricercatrice senior ed esperta di relazioni internazionali.

Adriano Iaria lavora nell'area Diplomazia Umanitaria di Croce Rossa Italiana come Humanitarian Advocacy Officer ed è FacultyMember al Master di II livello in Intelligence e Sicurezza della Link Campus University. Esperto in disarmo e diritto internazionale umanitario, collabora da docente e analista con diversi think tank italiani. Collabora con l'IRIAD dal 2013.

Juan Carlos Rossi, laureato presso La Sapienza di Roma in Scienze Politiche e Relazioni Internazionali, nel 2015 ha conseguito il Master di II livello in Tutela dei Diritti Umani presso la medesima università. Collabora con l'IRIAD dal 2015.

Maurizio Simoncelli, vicepresidente e cofondatore dell'IRIAD, storico ed esperto di geopolitica, collabora come docente di "Geopolitica, commercio delle armi e disarmo" presso il Master *"Nuovi orizzonti di cooperazione e diritto internazionale"* della FOCSIV/Pontificia Università Lateranense ed è direttore editoriale del mensile on line "IRIAD REVIEW. Studi sulla pace e sui conflitti".

Gian Piero Siroli, fisico sub-nucleare, ricercatore presso il Dipartimento di Fisica e Astronomia dell'Università di Bologna, svolge attività di ricerca in Fisica delle Particelle presso il CERN di Ginevra (occupandosi di sicurezza informatica) ed è associato anche all'Istituto Nazionale di Fisica Nucleare (INFN). Membro USPID, coordina le attività Pugwash nell'ambito della sicurezza cibernetica, anche in collaborazione con l'ONU ed altre istituzioni internazionali.

Guglielmo Tamburrini (PhD 1987, Columbia University, New York) è professore ordinario di filosofia della scienza e della tecnologia presso l'Università di Napoli Federico II. È stato coordinatore del primo progetto europeo sull'etica dei sistemi

robotici e dell'IA (ETHICBOTS, 2005-08). È membro dell'ICRAC (International Committee for Robot Arms Control) e dell'USPID.

ARCHIVIO DISARMO

L'Istituto di Ricerche Internazionali Archivio Disarmo è un'associazione di promozione sociale senza fini di lucro fondata a Roma nel 1982.In collegamento con diversi organismi nazionali ed esteri, favorisce la conoscenza scientifica e la divulgazione dei seguenti argomenti: sicurezza, disarmo, controllo degli armamenti, tutela diritti umani, gestione e risoluzione nonviolenta dei conflitti.

USPID

L'Unione degli Scienziati Per Il Disarmo ONLUS (USPID) è un'associazione di scienziati e ricercatori costituita nel 1983 con l'obiettivo, nel quadro della responsabilità sociale degli scienziati, di fornire informazione ed analisi su controllo degli armamenti e disarmo, incluse quelle relative all'impatto ambientale ed ai costi umani dello sviluppo e della diffusione degli armamenti.

IRIAD Review Pubblicazioni Recenti

	IRIAD Review	
S. Doro	*L'Iran a quarant'anni dalla Rivoluzione. Quali prospettive per uno dei principali attori sulla scena mediorientale?*	*Dicembre 2019*
G. Cassano	*Insicurezza nucleare: la minaccia del cyberterrorismo*	
B. Giuliani	*Il Trattato sulla Non Proliferazione delle Armi Nucleari in vista della Conferenza di Revisione 2020: problemi e prospettive*	
S. Doro	*Minori e conflitti armati. Quanto è ancora diffuso nel mondo l'utilizzo dei bambini soldato?*	*Gennaio 2020*
D. M. Trimonti	*Le operazioni di pace nel 2018*	
N. Scarpat	*Epidemia di Coronavirus e armi da fuoco negli Stati Uniti*	*Febbraio 2020*
S. Doro	*La Turchia: un attore fondamentale non solo per l'equilibrio del Medio Oriente*	
D. Talento	*L'evoluzione storica della "questione Cipro" e le rivendicazioni turche nei suoi spazi marini: il caso della Saipem 12000*	
T. Rinaldi	*La "Belt and Road Iniziative" e il rafforzamento della posizione della Cina nel mondo*	*Marzo/Aprile 2020*
B. Giuliani	*Le nuove traiettorie della politica di difesa e sicurezza cinese*	
S. Adrianopoli	*2019: aumentano ancora le spese militari mondiali*	
B. Gallo	*La minaccia invisibile delle armi biologiche: uno scenario in evoluzione*	*Maggio 2020*
F. Battistelli	*Natura devastata, animali maltrattati, esseri umani in pericolo: che cosa c'entra il Coronavirus con gli allevamenti intensivi?*	
N. Labanca	*Armi chimiche in Italia: storia e cronaca*	
M. Simoncelli	*Covid-19 ed equilibri geopolitici*	
M. Simoncelli *F. Sparagna* *D. Foschi* *G.Trama*	*Dal commercio senza vincoli all'export del 2020: 30 anni di 185*	*Giugno 2020*