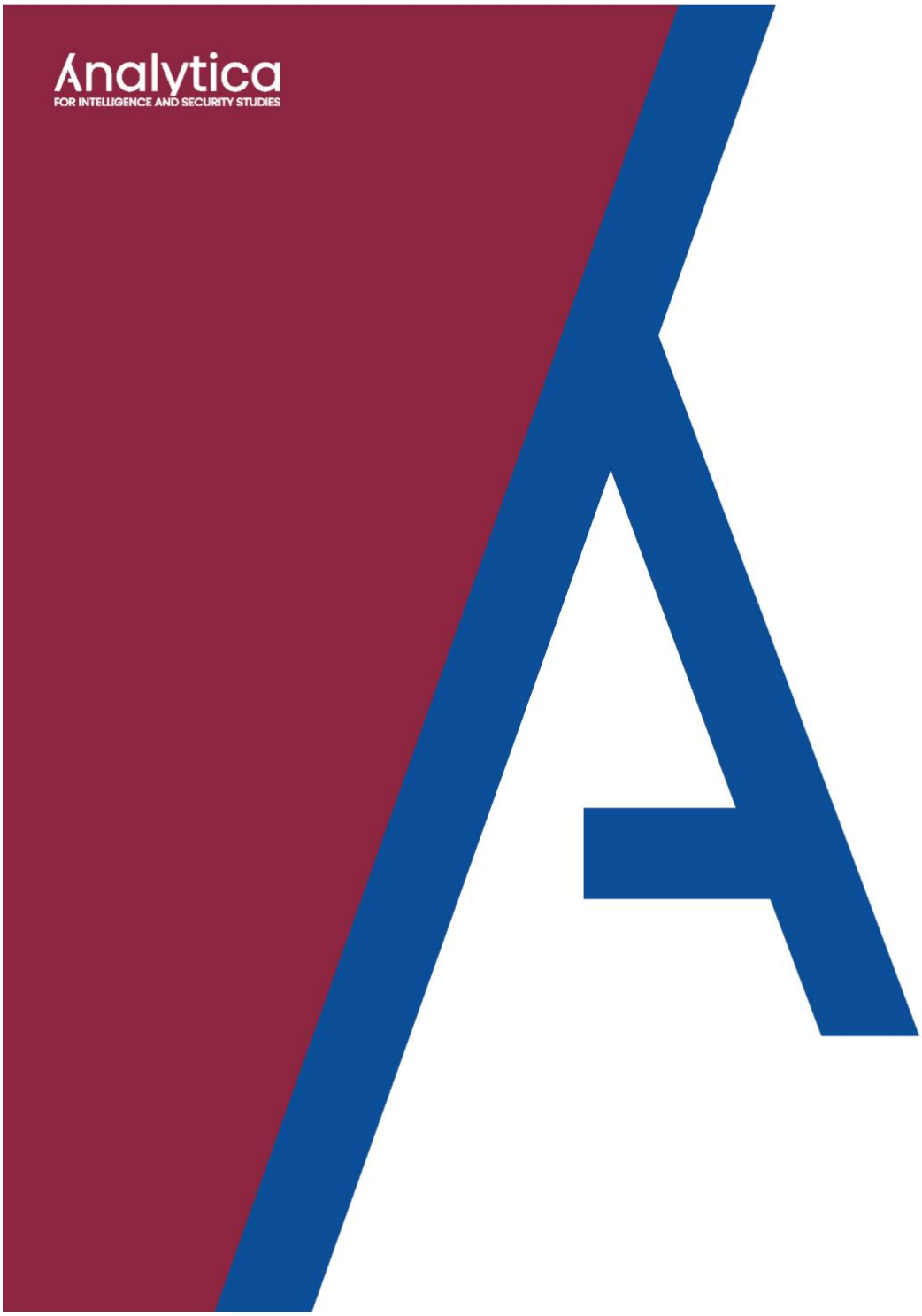

Nuove frontiere dell'Intelligenza Artificiale nel dominio cibernetico.

Cosimo Melella

Analytica for intelligence and security studies

Paper Cyber-Security

ISSN: 2784-8779

Nuove Frontiere dell'Intelligenza Artificiale nel mondo cibernetico.

Cosimo Melella

Correzioni e revisioni a cura del Dottor SPELTA Maurizio

Direttore del Dipartimento Cyber - Security

Torino, luglio 2021

Per “Intelligenza artificiale” o semplicemente IA si indica un insieme di processi tecnologici che hanno in comune la capacità di simulare i processi cognitivi umani, ossia la capacità di acquisire e apprendere nuove informazioni e concetti, elaborarli, strutturarli e conservarli in memoria. In un senso ancora più ampio l’intelligenza artificiale comprende un panorama diversificato di tecnologie e aree di sviluppo scientifico, dall’informatica alla matematica, alle neuroscienze. Sebbene il concetto d’Intelligenza Artificiale sia stato coniato per la prima volta negli anni ‘50, la capacità computazionale che favorisce in modo significativo le tecnologie necessarie alla c.d. Intelligenza Artificiale si è sviluppata significativamente solo negli ultimi anni e ora sta coinvolgendo numerosi aspetti delle società: dall’*Internet of Things* all’*automotive*, nella forma delle *self-driving cars*. Per evitare equivoci con i non addetti ai lavori, soprattutto nella prospettiva dei cambiamenti che le tecnologie e le progettazioni dell’Intelligenza Artificiale imporranno anche nei possibili conflitti, sarebbe però più corretto definire l’Intelligence Artificiale, come fanno già i ricercatori, “intelligenza aumentata”: ossia una IA declinata nella simbiosi uomo-macchina, dunque ben lontana dai cliché della fantascienza classica e semmai più vicina alle suggestioni alla *cyberpunk*[1].

Al di là delle seppur significative differenze, le tecnologie che identificano le aree di competenza riconducibili all’“intelligenza artificiale” possono essere suddivise in tre macro categorie: rilevamento e percezione, emozione e ragionamento e infine apprendimento automatico.

In particolare, nel dibattito contemporaneo, l’apprendimento automatico è la categoria scientifica che più spesso di altre viene usata come sinonimo d'Intelligenza Artificiale. In quest’ambito ci sono numerosi progressi che includono la capacità della macchina d’interpretare i dati e comprendere la semantica delle informazioni. L’apprendimento automatico è relativamente semplice da capire. Si tratta, infatti, di un insieme d' “input di dati” la cui elaborazione passa da un algoritmo che riesce a dedurre apprendendo da un determinato problema. In breve, le sofisticate tecniche d’IA odierne non mirano a superare le sfide computazionali di calcolo che richiederebbero computer sempre più potenti, piuttosto raccolgono i dati in modo più efficace per progettare processi migliori[2].

---

[1] A.M.A. Musleh Al-Sartawi, *Artificial Intelligence Systems and the Internet of Things in the Digital Era*, Berlino, 2021.

[2] A. Parisi, *Hands-On Artificial Intelligence for Cybersecurity: Implement smart AI systems for preventing cyber attacks and detecting threats and network anomalies*, Birmingham, 2019.

Intelligenza artificiale e cyberdefence.

L'Intelligenza Artificiale pone nuove sfide tanto ai soggetti privati quanto a quelli pubblici, come gli Stati; in particolare nel campo della *cyber defence* dove l'Intelligenza Artificiale è utilizzata sia per nell'attacco che nella difesa del dominio cibernetico. Nello specifico dell'alveo della difesa, l'intelligenza artificiale viene usata per implementare strategie di offesa e difesa più efficienti: lo sviluppo di strategie che tutelino risorse, reti, programmi e dati, da accessi non autorizzati che hanno lo scopo di manipolare, distruggere o comunque danneggiare in qualsiasi modo il *target* dell'attacco è proprio alla base della sicurezza informatica. Infatti, negli ultimi anni, l'adozione di tecniche d'intelligenza artificiale è aumentata significativamente, assumendo un ruolo cruciale nel rilevamento e nella prevenzione delle minacce informatiche.

Sistemi di sicurezza basati su tecnologie che "apprendono autonomamente" da attacchi informatici potrebbero imparare a rispondere più efficacemente a determinate azioni illecite aiutando a rilevare fin da subito le minacce in base a specifici comportamenti e/o attività su un'intera rete. Quindi, un sistema di sicurezza progettato con sistemi a "intelligenza aumentata", potrebbe sviluppare processi di monitoraggio del traffico mettendo in atto azioni correttive così da "normalizzare" i processi anticipatamente in base alle previsioni. Persino un *Intrusion Detection System*, un software utilizzato per identificare accessi non autorizzati ai computer o alle reti locali, se integrato da tecniche fondate sull'intelligenza artificiale affinerebbe le proprie tecniche di rilevamento delle intrusioni non autorizzate alla rete presidiata grazie a flessibilità, adattabilità, a calcoli e apprendimento rapido.

A causa dei notevoli progressi nelle tecnologie dell'informazione e della comunicazione, però, stanno emergendo anche nuove minacce. Tra queste ci sono gli attacchi di *phishing,* già noti per essere tra le principali minacce in rete. Per rispondere a queste minacce informatiche, sono state recentemente applicate varie tecniche innovative per sviluppare sistemi di filtro *antispam*. In particolare, in un recente studio Faris et al. hanno presentato un sistema di rilevamento e d'identificazione dello *spam e-mail* basato su un algoritmo genetico[3] e su una rete bayesiana ottenendo risultati notevoli in termini di accuratezza e precisione[4].

---

[3] Si tratta di un algoritmo che permette di valutare diverse soluzioni di partenza introducendo e ricombinando elementi di disordine, analogamente alla riproduzione biologica e alle mutazioni genetiche casuali, così da produrre nuove soluzioni che vengono poi selezionate e scelte nel tentativo di convergere verso soluzioni "di ottimo".

[4] A.M. Al-Zoubi, H. Faris, M. A. Hassonah, *Spam profiles detection on social networks using computational intelligence methods: The effect of the lingual context,* Volume: 47 issue: 1, page(s): 58-81, August 7, 2019.

Sono state poi proposte tecniche che utilizzano processi d'intelligenza artificiale volte a rilevare o classificare i *malware*, rilevando le intrusioni di rete. E se, come spesso si legge dalla stampa, i *threat actors* si concentrano sempre di più spesso su obiettivi come agenzie di sicurezza aziendale e organizzazioni governative, allora per difendersi sarà necessario sviluppare diversi processi che adoperino l'intelligenza artificiale e che applichino "alberi decisionali" per configurare *Intrusion Detection System* capaci di rilevare *Advanced Persistent Threat* (un sofisticato attacco informatico che, a prescindere dall'artefice della minaccia, utilizza tecniche avanzate per creare falle difficilmente rilevabili e non essere tracciato).

Questi sistemi possono rilevare le intrusioni fin dall'inizio e reagire rapidamente per ridurre i danni al minimo. I risultati empirici hanno mostrato che il sistema proposto ha raggiunto un alto tasso di successo. Infatti, in uno studio apposito, Sharma et al. hanno presentato un modello per il rilevamento delle *APT* basato su più classificazioni parallele. Si tratta di un nuovo modello che rappresenta una promettente base per i moderni sistemi di rilevamento delle intrusioni (gli *Intrusion Detection System*). A differenza di altri approcci, però, la tecnica DFA-AD per rilevare un attacco *APT* si basa su più classificatori paralleli: ogni metodo di classificazione si concentra sul rilevamento della tecnica di attacco delle *APT* in modo indipendente[5].

Intelligenza artificiale e cybercrime

Come già accennato in precedenza, se grazie all'intelligenza artificiale si stanno diffondendo nuove e innovative tecniche di difesa, d'altra parte stanno nascendo anche nuove tecniche di *cybercrime*, ossia di crimine informatico. L'intelligenza "aumentata", infatti, può essere utilizzata tanto come difesa quanto come arma per aumentare l'efficacia delle nuove generazioni di *malware* rendendoli più autonomi, più sofisticati, più veloci e più difficili da rilevare. Questi nuovi "programmi malevoli intelligenti" programmati da una nuova generazione di *cybercriminali* riuscirebbero ad auto propagarsi in una rete o in un sistema informatico sulla base di una sequenza di decisioni autonome, adattate in modo intelligente ai parametri del sistema che il *malware* ha deciso d'infettare.

Inoltre, il *malware* potrebbe adattarsi all'ambiente o utilizzare le conoscenze acquisite durante uno specifico attacco per inviare autonomamente informazioni alla stazione di Comando e Controllo (C&C) o addirittura modellare attacchi adattivi indipendenti sfruttando i dati acquisiti e sviluppando degli attacchi ai sistemi informatici "cuciti su misura". Uno degli obiettivi finali del

[5] P.K. Sharma,, S.Y. Moon, D. Moon et al. DFA-AD: *a distributed framework architecture for the detection of advanced persistent threats. Cluster Comput* 20, 597–609 (2017). https://doi.org/10.1007/s10586-016-0716-0.

malware sarebbe infine nascondere la propria presenza e i propri scopi per evitare di essere rilevato dai sistemi di sicurezza *anti-malware*.

Parallelamente allo sviluppo dei *malware*, sono inoltre in corso studi per l'applicazione di tecniche d'ispirazione biologica. Ad esempio, Ney et al. hanno presentato come compromettere una rete codificando il malware come se fosse una sequenza di *DNA*. Successivamente hanno delineato uno "sciame di malware" come sfondo per un futuro sistema anti-malware. Più precisamente, il prototipo di questa tipologia di virus ha simulato il comportamento di una organizzazione di tipo *swarm* (un sistema che ha la capacità di auto organizzarsi e adattarsi) in cui le sue informazioni sono state archiviate e visualizzate sotto forma di rete complessa[6].

Per le istituzioni nazionali che si occupano di difesa e sicurezza, l'emersione di nuove forme di offesa che fanno riferimento a tecniche legate all'intelligenza artificiale è preoccupante perché esistono già applicazioni anche se a livello di prototipi (come ad esempio nella guerra cognitiva, un sotto insieme della *cyberwarfare*, dove – semplificando – è la mente umana a rappresentare il nuovo campo di battaglia dove vengono impiegate e integrate capacità informatiche, psicologiche e d'ingegneria sociale). Gli attacchi informatici basati sull'intelligenza artificiale fanno presagire la personalizzazione e la manipolazione dei sistemi sia informatici che sociali (con possibili effetti cinetici e nascita di nuovi teatri di scontro fisico), introducendo il rischio derivante dal trasferimento di competenze tecniche da un *hacker* a un algoritmo. L'aumento delle difese potenziate dall'intelligenza artificiale, incorporate nei sistemi di difesa nazionali, potrebbe comunque presentare vulnerabilità agli attacchi che prevedono, manipolano e sovvertono la funzionalità degli algoritmi difensivi.

In tutto il mondo pochi attori prestano altrettanta attenzione quanto le forze armate sui possibili impatti che i sistemi d'intelligenza artificiale avranno nell'ambito della sicurezza. In particolare i leader politici e i vertici militari di alcuni paesi NATO (quali USA, UK, Francia ed Estonia) hanno suggerito d'investire in progetti relativi all'intelligenza artificiale per avvantaggiarsi sugli avversari geopolitici. L'aspettativa è che i processi d'IA porteranno a un'inevitabile trasformazione e alterazione delle relazioni tra Stati in termini sia strategici che operativi.

In particolare, sebbene ci sia un piccolo ma crescente gruppo di lavoro nell'*Intelligence*

[6] Xuejing Zhao, Chen Wang, Jinxia Su, Jianzhou Wang, *Research and application based on the swarm intelligence algorithm and artificial intelligence for wind farm decision system,* 2018. https://www.researchgate.net/profile/Xuejing-Zhao/publication/329098989_Research_and_Application_Based_on_the_Swarm_Intelligence_Algorithm_and_Artificial_Intelligence_for_Wind_Farm_Decision_System/links/5bff71a1a6fdcc1b8d4a0a37/Research-and-Application-Based-on-the-Swarm-Intelligence-Algorithm-and-Artificial-Intelligence-for-Wind-Farm-Decision-System.pdf.

Community statunitense che coinvolge agenzie come *NSA* e *DHS* sulla capacità dell'IA d'influenzare la propaganda politica e non solo, la reportistica che affronta il dibattito sui possibili sviluppi futuri dell'IA nell'ambito della disinformazione online è ancora scarna[7].

Inoltre, il lavoro in corso tende a coinvolgere solo analisi descrittive di scenari di minaccia, senza considerare come l'aumento delle capacità *cyber,* in particolare l'applicazione di tecniche di *machine learning*, alteri le dinamiche strategiche.

Infatti, mentre le nuove tecniche di apprendimento dell'avversario sembrano destinate a migliorare il *kit* di strumenti sia dei difensori che degli attaccanti, con l'Intelligenza Artificiale probabilmente saranno questi ultimi a risultare in vantaggio. D'altra parte se la deterrenza tattica è incredibilmente difficile da raggiungere, un equilibrio tra le tecniche difensive e quelle offensive sarebbe solo momentaneo ed effimero e non potrebbe essere mantenuto. Infatti, le nuove capacità di apprendimento che vengono impiegate su larga scala in operazioni di *routine* aggiungono ulteriore complessità ai processi che operatori e analisti d'*intelligence* devono considerare nella "difesa del perimetro".

Nuove tecniche offensive.

Analizziamo ora in che modo l'IA possa implementare o aggiornare le operazioni informatiche offensive: l'IA, ovvero l'apprendimento automatico, riduce l'efficacia delle misure difensive convenzionali e rende gli attacchi potenti più accessibili anche ad attori di medio livello. Questo in quanto l'intelligenza artificiale riesce a sviluppare adattabilità, velocità e opportunità di programmazione senza precedenti a costi relativamente contenuti e accessibili.

La diffusione dei processi legati all'IA fa presagire minacce ben più gravi di quelle affrontate finora sia per quanto riguarda la loro rilevanza strategica sia per le dimensioni delle superfici dei sistemi.

Vanno considerati due livelli che aumentano la pericolosità: il primo prende in esame le occasioni che consentirebbero a un *malware* di sfruttare i dati ottenuti tramite l'infezione in modo da valutare dove e quando potrebbe verificarsi un'ulteriore infezione; il secondo livello, invece, riguarda le dinamiche che impiegano l'intelligenza artificiale programmata sulla capacità dei *malware* di selezionare opzioni per un'ulteriore diffusione in modo indipendente dai *target* di partenza. Nello specifico, una volta sovrascritto un programma di avvio con la funzione di testa di ponte, in pochi minuti, il *malware* prende di mira e compromette le macchine aggiuntive, senza

[7] C. Cunningham, *Cyber Warfare – Truth, Tactics, and Strategies: Strategic concepts and truths to help you and your organization survive on the battleground of cyber warfare*, Birmingham, 2020.

seguire uno schema di *targeting* specifico[8].

In questo modo il codice malevolo non solo è in grado di calibrare il proprio attacco a una velocità specifica, ma può anche selezionare le vittime sulla base di un'analisi "intelligente" del potenziale successo. Il *malware* è semplicemente programmato per intraprendere un'azione più attenta e *Trickbot* rappresenta un esempio di come *malware* simili potrebbero funzionare e in cui, attraverso una rapida comprensione della superficie di attacco, si giunge a un'articolata strategia d'infezione. Un altro esempio di un'analisi più accurata della superficie di attacco è quella che prende in considerazione la ricchezza di dati e *metadati* ottenibili attraverso la tradizionale *information gathering*. In futuro potrebbero benissimo esserci campagne informatiche di natura criminale o politica che abbiano come target la ricchezza dei dati messe a disposizione degli aggressori per l'analisi.

Il *malware* che riuscisse a penetrare nelle difese potrebbe intraprendere analisi ambientali e determinare quale processo sia più adatto per attaccare nuovi *target*. In questa forma l'attacco informatico intrapreso grazie alle tecniche d'Intelligenza Artificiale non è molto diverso da quello condotto da software generalmente già impiegati da alcune da alcune *APT* (qualcuna potrebbe addirittura essere sponsorizzata addirittura da qualche Stato). Si tratta sostanzialmente di un'abilità più accessibile che consente a *script kiddies,* cioè a individui che non sono in possesso delle abilità necessarie a sviluppare programmi o *exploit* sofisticati, di usare strumenti intelligenti per adattare i *toolkit* di attacco a propria disposizione per diverse superfici di attacco.

Naturalmente, se il potenziale dell'intelligenza artificiale può essere riassunto in una maggiore adattabilità e furtività dei comportamenti dannosi, si può dire lo stesso anche per il potenziale che potrebbero avere le difese informatiche abilitate dall'intelligenza artificiale. Inoltre, sarebbe ingiusto affrontare qualsiasi discussione sul potenziale impatto dell'IA sui conflitti informatici, senza considerare che le nuove tecniche di apprendimento e rilevamento sosterranno anche significativi sforzi difensivi tanto che i processi fondati sulle IA diventeranno indispensabili nelle difese perimetrali convenzionali.

Ciò che è particolarmente unico nell'intersezione tra IA e i processi del conflitto cibernetico è la centralità assunta dai processi extra dominio rispetto alle operazioni all'interno del dominio, ossia sono incentivate le operazioni che hanno effetti al di là dei risultati raggiunti all'interno del dominio *cyber*. Al di là di quasi tutte le altre implicazioni, l'aggiornamento delle tecniche informatiche "convenzionali" fa presagire una nuova sfida per la strategia IT abbastanza semplice da affrontare. In particolare, si avrà uno minore spazio di ottimizzazione tra costi/benefici e gli

[8] A. Parisi, *Hands-On Artificial Intelligence for Cybersecurity: Implement smart AI systems for preventing cyber attacks and detecting threats and network anomalies*, cit.

attaccanti potrebbero tentare di addossare ai difensori i costi dell'attacco stesso. In altre parole, se esistono strumenti intelligenti che possano evitare il rilevamento in modo affidabile ed efficace, è altrettanto probabile che verranno sviluppati sistemi capaci di sfruttare percorsi "secondari" verso i bersagli. In conclusione è plausibile sostenere che la proliferazione di attacchi attraverso azioni laterali per ottenere effetti nel dominio digitale porterà a una maggiore frequenza d'incidenti in aree in cui in precedenza si pensava che la minaccia fosse stata contrastata.

Nuove strategie difensive.

Fino a oggi, le tradizionali operazioni di deterrenza e approcci diplomatici convenzionali hanno dato prova di essere dei solidi pilastri della politica estera informatica degli Stati della NATO, in particolare degli USA; tuttavia, il successo di una politica della deterrenza sembra possibile solo laddove esista un preciso allineamento situazionale con altri sforzi, a causa della variabilità delle condizioni che consentono le azioni cibernetiche e nelle reazioni delle complesse infrastrutture militari e civili statali[9].

Un'ultima considerazione sull'IA è il modo in cui gli sforzi per proteggere oggi il cyberspazio potrebbero precipitare. Come osserva Healey: "La difesa in avanti si basa naturalmente su una grande fiducia tra gli alleati e il settore privato. Tuttavia, le azioni implicite nella strategia sono inevitabilmente tra le più invasive. Questo produce una sfida di fiducia per il successo della strategia, senza soluzioni facili e molte linee di faglia in cui l'*escalation* non è solo possibile ma probabile". Il decisore politico e il settore privato devono fare i conti con un dominio, il cyberspazio, che diventerà il principale terreno di scontro per interferire nelle dinamiche politiche di un paese, destabilizzandolo con attacchi d'influenza cibernetica[10].

In generale, l'applicazione dell'IA nei *toolkits* funzionali degli apparati di sicurezza nazionale (per difesa e attacco) genera una tensione continua nella condotta delle operazioni in corso. L'esistenza di robusti sistemi d'IA da parte degli Stati NATO genera un problema di apprendimento: più gli avversari cercheranno di comprendere e superare le strategie difensive innovative più gli apparati di sicurezza dovranno lavorare per modellare il comportamento delle amministrazioni di cui sono parte per offrire una risposta efficacie. Inoltre, dato l'incentivo a utilizzare *software* abilitati con l''IA al *track record* (cioè alla capacità di tracciare informazioni di

[9] A. Parisi, *Hands-On Artificial Intelligence for Cybersecurity: Implement smart AI systems for preventing cyber attacks and detecting threats and network anomalies*, cit.
[10] J. Healey K. T. Jordan, *NATO's Cyber Capabilities: Yesterday, Today, and Tomorrow*, Atlantic Council, 2014 https://www.jstor.org/stable/resrep03426.

un individuo o di un gruppo), è probabile che l'esistenza di tali sistemi ostacoli lo sviluppo di regole di condotta chiare e definite

Un'ultima riflessione sembra particolarmente degna di essere presa in considerazione: l'errata convinzione che un attacco informatico sia un episodio occasionale e non l'indizio di un conflitto in atto. La mancanza di consapevolezza può produrre gravi conseguenze concatenate difficilmente controllabili. Sebbene gli attacchi automatici oggi rappresentino una sfida unica in cui poter mettere in campo risposte automatiche difensive, gli effetti a cascata a un certo punto tenderebbero a cessare semplicemente a causa di *backstop* nell'algoritmo.

Quello che dunque potrebbe sembrare un attacco potrebbe essere invece uno sforzo per studiare lo spazio di battaglia o saggiare le capacità del nemico in attività non belliche. Al di là di questo livello di discussione, tuttavia, gli studiosi hanno prestato anche attenzione all'ipotesi di effetti a cascata con *escalation* nello spazio cinetico: infatti con l'intelligenza artificiale esiste il rischio concreto che più interazioni (costanti del tipo *back and forth*) possano produrre una massa critica di attività tali da generare effetti negativi significativi anche nel mondo non digitale.

Un esempio comunemente citato, tenendo presente le dovute differenze di un evento di massa critica, è il *flash crash* del mercato azionario del 6 maggio 2010. Tuttavia, nonostante sia stato ampiamente studiato, non si è giunti a concordare su quale possa essere stata la causa: convenzionalmente si attribuisce la perdita al *Dow Jones* di quasi 1.000 punti in soli 36 minuti alla reazione a un insolito disturbo del mercato da parte degli algoritmi automatizzati. La ragione di una perdita di trilioni di dollari nel mercato (anche se è venne assorbita rapidamente nelle ore successive) sarebbe dunque una vendita accidentale di alcuni ordini di grandezza superiore a quanto ci si aspetterebbe. Guardando l'evento, è facile immaginare cosa potrebbe causare una "battaglia" tra AI, ossia tra algoritmi automatizzati più o meno elementari: gli effetti potrebbero variare dalla sospensione delle funzioni d'infrastrutture critiche essenziali fino a determinare una controffensiva informatica tale da sollecitare una risposta fisica all'attacco nel dominio cibernetico.

In generale, le tecniche di apprendimento dell'Intelligenza Artificiale aggiungono complessità a condizioni operative articolate e multiformi già attive nel cyberspazio e possono contribuire a un'escalation del comportamento offensivo.

Ciò che più dovrebbe preoccupare, e non solo nei processi di conflitto informatico, è che gli avversari, grazie a un maggior livello di sofisticazione, potrebbero lanciare azioni offensive *online* per ottenere effetti negli altri domini e in particolare in quello fisico. Le implicazioni per le attuali

strategie di conflitto cibernetico, principalmente dette di difesa dai paesi NATO (in linea con quanto stabilito a Varsavia nel 2016), sono numerose e devono ancora essere valutate a fondo man mano che la letteratura sull'argomento si svilupperà in futuro[11].

Tuttavia, alcuni suggerimenti immediati sono evidenti: i *decision makers* devono innanzitutto riconoscere che ci sono più *layers* di sfida nell'aumentare le schermaglie cibernetiche con tecniche d'intelligenza aumentata. L'IA riduce le possibilità di deterrenza aumentando la concorrenza degli Stati nel cyberspazio in un gioco per tutti tutt'altro che a somma zero o a somma positiva (anzi, nel *long run* in termini di teoria dei giochi potrebbe essere un gioco a somma negativa).

L'IA intensifica e aggiunge una nuova dimensione alle sfide di attribuzione nelle operazioni informatiche. In poche parole, date le maggiori opportunità d'ingaggio grazie ai nuovi modelli d'intelligenza artificiale che presto saranno ampiamente diffusi fra le agenzie di sicurezza, anche per i c.d. *Rogue States*, la vera sfida sarà rimanere costantemente informati sull'integrità dei sistemi difensivi. Inoltre, l'Intelligenza Artificiale intensifica e aggiunge una nuova dimensione alle sfide di validità e attribuzione nelle operazioni informatiche. Il successo nell'affrontare le sfide dell'implementazione dell'intelligenza artificiale per scopi di sicurezza nazionale, dipenderà probabilmente dall'approccio adottato dalle organizzazioni: tanto nel fidarsi dei propri sistemi d'intelligenza artificiale quanto nel gestire l'interazione tra operatori umani e macchine[12].

Infine, sembra chiaro che lo sviluppo, la valutazione e la convalida della strategia devono derivare da un'attenta analisi e comprensione di quali potrebbero essere gli scopi e le motivazioni strategiche degli avversari. Come alcuni studiosi sostengono da tempo in termini sia impliciti che espliciti, faremmo bene a formulare le nostre analisi in termini di logica dei processi di conflitto non cibernetico e, in particolare, in termini d'IA.

Conclusioni.

Dal 1974, ossia da quando è stato lanciato il primo attacco di *Denial of Service* (*DoS*), la

[11] C. Cunningham, Cyber Warfare – Truth, Tactics, and Strategies: Strategic concepts and truths to help you and your organization survive on the battleground of cyber warfare, cit.
[12] B. Valeriano, B. Jensen, R. C. Maness, *Cyber Strategy: The Evolving Character of Power and Coercion*, Oxford University Press, 2020.

complessità, il numero e l'impatto degli attacchi informatici sono aumentati in modo esponenziale e, ovviamente, man mano che gli attacchi informatici sono diventati più mirati e potenti lo sono diventate anche le contromisure. Mentre all'inizio i primi strumenti di sicurezza si limitavano a rilevare le firme dei virus e a prevenire la loro esecuzione, oggi troviamo soluzioni progettate per fornire una protezione contro un'ampia gamma di tipi di attacchi e vari sistemi di destinazione; tuttavia, è diventato sempre più difficile proteggere le risorse informative nel mondo virtuale.

I sistemi di sicurezza devono adattarsi costantemente ai cambiamenti degli ambienti, delle minacce e degli attori coinvolti nel gioco informatico. La realtà informatica, invece, appare in qualche modo diversa: le metodologie di difesa vengono regolarmente adattare agli attacchi noti ma a causa della mancanza di flessibilità e robustezza, i sistemi di sicurezza generalmente non sono in grado di adattarsi automaticamente ai cambiamenti che avvengono nell'ambiente circostante. Le tecniche d'Intelligenza Artificiale, a causa della propria flessibilità e adattabilità possono aiutare a superare varie carenze degli strumenti di sicurezza informatica odierni. Sebbene l'Intelligenza Artificiale abbia già notevolmente migliorato la sicurezza informatica, ci sono ancora serie preoccupazioni. Studiosi ed esperti hanno espresso allarme e preoccupazioni di tipo etico per il ruolo crescente che l'IA svolge nel cyberspazio.

Sebbene la consapevolezza delle minacce informatiche sia aumentata e siano state investite ingenti somme di denaro e siano stati compiuti sforzi per combattere la criminalità informatica, la capacità delle organizzazioni legali di proteggere le proprie risorse virtuali non è ancora sufficiente. Le parti coinvolte nel cyberspazio vanno da singoli utenti a organizzazioni private e attori non statali: per questo le fonti delle minacce informatiche sono molteplici.

Ricapitolando, la maggior parte degli attacchi informatici segue determinate fasi di attacco che possono essere descritte attraverso una metrica specifica, la *Cyber Kill Chain.* Supponendo che ogni sequenza di attacco inizi con una fase di ricognizione, in cui un *threat actor* cerca di comprendere le vulnerabilità di un sistema attaccato, la fase successiva sarà quella di preparazione, durante la quale i punti deboli scoperti saranno utilizzati per sviluppare codice malevolo, a cui seguirà la fase in cui il malware verrà inviato al potenziale bersaglio. Con la consegna, avvenuta con successo, si verifica l'*exploitation*, ossia il momento durante il quale il malware attiva l'installazione del codice.

Successivamente, il sistema compromesso consentirà la creazione di una stazione di comando e controllo (*C&C*) in modo che l'aggressore possa avviare azioni dannose, e un utente malintenzionato possa cercare di comprendere le vulnerabilità di un sistema attaccato.

Gli attacchi informatici guidati dall'intelligenza artificiale differiscono in modo significativo dalle minacce digitali più convenzionali che hanno occupato professionisti e ricercatori negli ultimi tre decenni. È anche possibile, anzi probabile, che gli effetti si manifesteranno soprattutto al di fuori del cyberspazio. Tuttavia, la centralità del dominio *cyber* per l'implementazione e il funzionamento di sistemi d'intelligenza artificiale proporrà nuovi scenari operativi all'interno del quinto dominio stesso, mettendo in discussione diversi presupposti delle attuali strategie di prevenzione dei conflitti informatici.

I notevoli progressi nella tecnologia dell'informazione hanno portato all'emersione di nuove sfide per la sicurezza informatica. Strategicamente, le operazioni informatiche condotte dall'intelligenza artificiale potrebbero essere in grado di perseguire altri obiettivi oltre quelli programmati dagli avversari. Forse l'aspetto più preoccupante è che la centralità d'Internet, soprattutto per i nuovi sistemi d'intelligenza artificiale incorporati nelle aree della sicurezza nazionale, dimostra che avversari sofisticati potrebbero essere incentivati a lanciare azioni offensive online per ottenere effetti negli altri domini (terrestre, marittimo, aeronautico, spaziale).

Questi scenari dovrebbero essere fonte di notevole preoccupazione per i *decision makers*: infatti, una tale capacità di elaborazione informatica in un ambiente globale eterogeneo creerà uno spazio di battaglia ancora più contorto di quello tuttora esistente e introduce nuove sfide per la difesa su larga scala e amplifica alcuni rischi, come la possibilità di escalation in domini diversi da quello *cyber*.