WHITEPAPER

ATTACCHI INFORMATICI

COME SFRUTTARE L'INTELLIGENZA ARTIFICIALE E I BIG DATA PER UNA SICUREZZA A 360 GRADI

CYBEROO

.ABSTRACT

L'intensificazione degli **attacchi informatici** è dovuta a diversi fattori concomitanti, che sono presenti nelle organizzazioni, viste come potenziali target di attacco. Uno dei fattori che sta contribuendo a facilitare i **cybercriminali** è l'utilizzo di **algoritmi di Intelligenza Artificiale** (AI), che aumentano e specializzano le capacità e il successo degli attacchi.

Anche i **cyberdefender,** deputati a proteggere e prevenire, possono adottare sistemi e mezzi di **AI**, quali per esempio il **machine learning** e le **tecniche deduttive dei sistemi inferenziali**, oppure sistemi di big data che implementano capacità di analisi avanzata. Quando le due tecnologie sono fuse insieme si possono ottenere sistemi di "**situational awareness**" e di valutazione prospettica basate su scenari "**what if**", per abilitare la difesa e la protezione a fronte di dati molto probabili e prossimi alla certezza.

Nel White Paper sono analizzati gli ambiti tecnologici della **AI** e dei **Big Data**, prima come trend tecnologici, poi singolarmente e infine come soluzione integrata. In particolare, sono descritti gli ambiti di utilizzo della AI e dei Big Data nei diversi settori di mercato e nell'ICT per una panoramica introduttiva.

Entrando nel merito della sicurezza informatica, si specializza la trattazione sulle modalità con cui gli algoritmi di Intelligenza artificiale e quelli che adottano Big data possano singolarmente abilitare capacità di difesa preventiva e di detection senza ritardi.

Infine, si spiega come rendere più efficaci ed efficienti queste due tecnologie mediante l'integrazione in un'unica **soluzione** evoluta di difesa "Intelligent Driven" e basata sull'esperienza, che possa fornire analisi e valutazioni ad ampio spettro.

.INTRODUZIONE

Gli **attacchi informatici** mostrano un **trend crescente anno dopo anno**, favoriti da una scarsa adozione delle prassi basilari di difesa e da approcci di protezione ancora non completamente appropriati. Le cause principali da parte delle organizzazioni sono imputabili all'insufficienza di investimenti in cyber security, la crescita continua di estensione e complessità della superficie di attacco, la debolezza della supply chain e la cronica mancanza di una cultura diffusa verso la sicurezza informatica. Tuttavia, dal punto di vista degli avversari, il motivo principale degli attacchi resta il **cybercrime** ovvero gli **alti guadagni** che i criminali informatici possono trarre a fronte di spese relativamente contenute per dotarsi degli strumenti necessari.

Un ulteriore elemento di complicazione è costituito dall'adozione da parte dei criminali informatici di **tecnologie innovative**, che permettono l'automazione degli attacchi, la specializzazione per dominio di mercato, la profilazione e la targetizzazione delle vittime. Le prime due capacità sono implementate mediante sistemi esperti, bot e sistemi di machine learning che costituiscono applicazioni della AI. Le altre due possono essere implementate mediante adozione di analisi su Big Data.

Naturalmente queste stesse tecnologie utilizzate dai cyberdefender possono costituire valide soluzioni per aumentare le capacità di **prevenzione e di difesa**, mediante lo **studio e la profilazione degli avversari**, delle loro **tecniche di attacco** e degli **strumenti di attacco preferiti**, in modo da implementare **difese ad hoc**.

Fino a oggi **AI e Big Data** sono state utilizzate in molti contesti dell'Information and Communication Technology (ICT) per potenziare, specializzare e in generale migliorare le prestazioni e i risultati. Ulteriormente e singolarmente specializzate alla cyber security sono in grado di **abilitare difese più efficienti** nell'incident management, nel crisis management, ma anche nella difesa preventiva, grazie alle analisi degli scenari di attacco a cui prepararsi.

1.TREND NELL'ADOZIONE DI INTELLIGENZA ARTIFICIALE E BIG DATA

Dalla sua nascita, il 1956, a oggi l'Intelligenza Artificiale (AI) ha raggiunto notevoli traguardi in diversi campi di applicazione. Definita come "quel un ramo dell'informatica che permette la programmazione e progettazione di sistemi sia hardware che software per dotare le macchine di caratteristiche considerate tipicamente umane" (fonte: http://www.intelligenzaartificiale.it/), la AI è stata oggetto di ricerca sperimentazione e utilizzo in **campo biologico** (anni '60), **nella teoria dei giochi** (scacchi, Otello, GO, backgammon, etc.), **nella teoria dei linguaggi formali e delle decisioni**.

Sebbene l'utilizzo dell'AI per la simulazione ed elaborazione del linguaggio naturale, l'apprendimento e la soluzione di problematiche vicine all'uomo sia iniziata dalla seconda metà degli anni '60, l'implementazione pratica è stata progressivamente crescente e commisurata alle effettive capacità operative dei sistemi intelligenti.

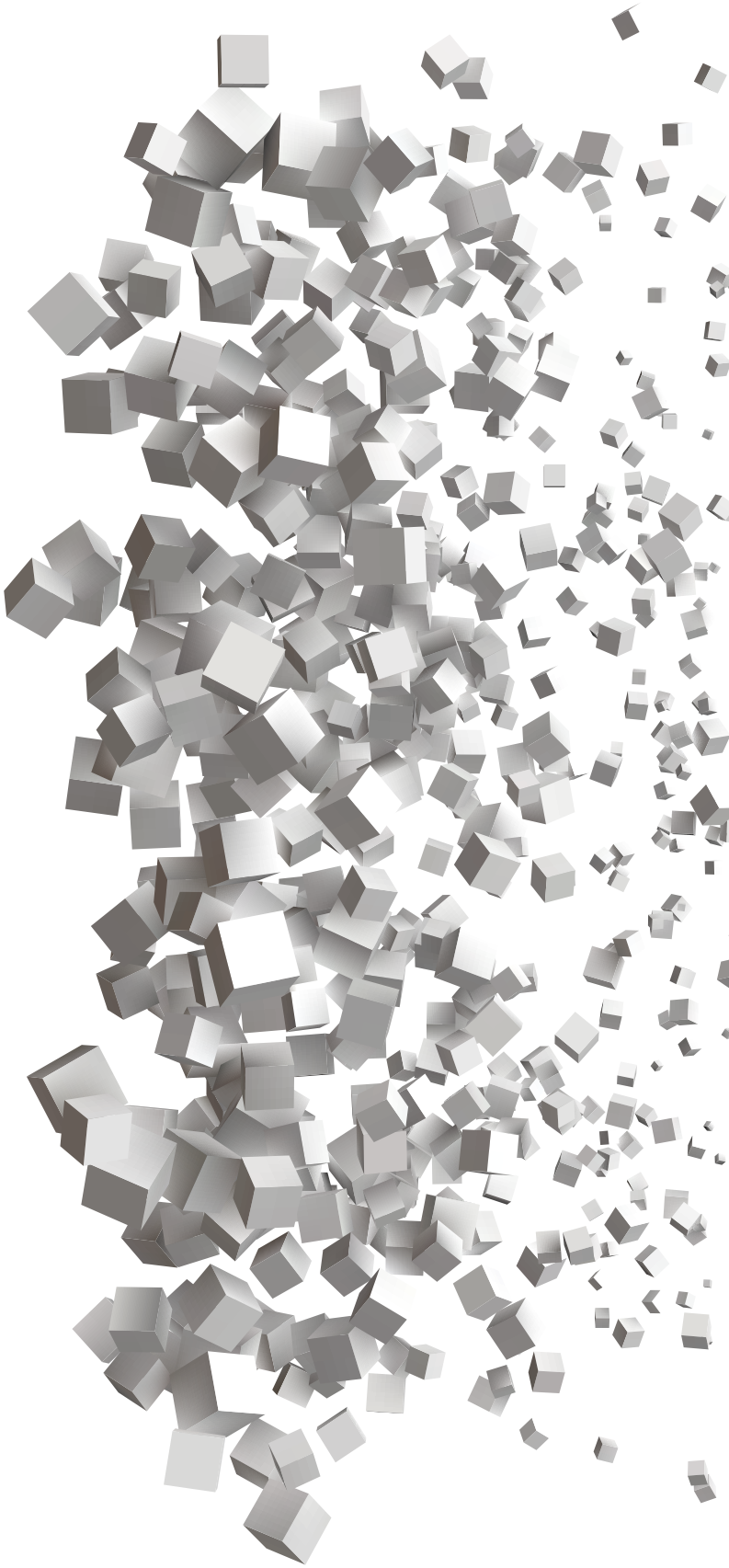
Oggi l'apprendimento e il ragionamento, sviluppati mediante utilizzo di reti neurali e algoritmi in grado di riprodurre ragionamenti tipici degli esseri umani, consentono di "**simulare la presa di decisioni**", effettuando scelte a seconda dei "contesti" in cui sono inseriti. I "contesti" a loro volta, per essere elaborabili hanno avuto bisogno di essere sviluppati in una **sistema di rappresentazione della conoscenza** (quella di base e di tipo allargato o specializzato con riferimento a un ambito, n.d.r).

Un decisivo passo in avanti è avvenuto con il **Machine Learning** (apprendimento automatico, n.d.r.) ovvero lo sviluppo di algoritmi in grado di imparare dai propri errori e quindi capaci di migliorare il comportamento della macchina che riesce a simulare il modo umano di "fare esperienza".

Tutte queste capacità di deduzione e decisione basate sull'apprendimento sono state applicate al contesto **automobilistico**, per la guida autonoma, per l'elaborazione di percorsi ottimizzati in funzione di tempo traffico e del tipo di strada e per realizzare sistemi di cambi di velocità in auto a guida semi-autonoma (implementazione di logica fuzzy, n.d.r.). In **ambito medico**, l'uso di reti neurali consente di ottimizzare le analisi del battito cardiaco, o la diagnosi di alcune forme tumorali e più recentemente è stata avviata la realizzazione di robot di accompagnamento.

In **campo high tech** sono innumerevoli gli **usi dei cosiddetti bot**, ossia sistemi di AI specializzati per funzioni specifiche: nel **customer care** per realizzare sistemi automatici di risposta e interazione con gli utenti, per la **domotica**, con gli assistenti intelligenti, per i PC e mobile per realizzare **assistenti interattivi**, o per realizzare app che funzionino come veri e propri sensori, modificando le impostazioni di comfort e sicurezza del device per adattarsi alla situazione. Anche l'**industria 4.0** pianifica di adottare sistemi esperti e dotati di machine learning per ottimizzare i processi produttivi.

Più recente rispetto alla AI, il concetto di **Big Data** ha, invece, una origine che si assesta alla fine degli anni '90, ma che ha visto poi un suo utilizzo massivo e iperbolico circa 10 anni fa, quando le big tech **Amazon, Facebook e Google** si ritrovarono a dover gestire quantitativi enormi di dati e investirono per trarre vantaggio da tutti questi dati aggregati. Se i dati sono una raccolta di fatti, come valori o misurazioni e sono spesso il risultato della combinazione, del confronto e dell'esecuzione di calcoli sui dati, i "Big Data sono risorse informative ad alto volume, ad alta velocità e ad alta varietà che richiedono forme innovative e convenienti di elaborazione delle informazioni e consentono approfondimenti, processi decisionali e automazione dei processi" (fonte: Glossario Gartner).



L'utilizzo più noto è la capacità di profilazione dell'utenza a fini del **marketing targettizzato e specializzato sui gusti degli individui**, ma lo spettro di usi meno nobili è possibile: controlli sulla salute per conto di società assicurative o datori di lavoro, o da società di recruiting, ma anche controllo della popolazione. Da cui derivano problematiche di privacy e di tutela dei dati personali. Quale che sia l'uso finale, i **Big Data** sono soggetti a **due categorie principali di azioni: il Management** ovvero la serie di processi che riguardano l'acquisizione e la memorizzazione delle informazioni di grandi dataset, attività che richiede accorgimenti specifici e **l'Analytics**, l'analisi di questi dati che, grazie ad appositi algoritmi di modellazione, avviene nel modo più rapido possibile ed adeguato alle performance delle aziende. "Gli algortitmi per i Big Data vengono creati per poter consentire uno studio del flusso dei dati, e per questa ragione che devono essere parametrici, multilivello e precisi" (fonte: http://www.intelligenzaartificiale.it/big-data/).

Normalmente i Big Data sono utilizzati per la profilazione nel **campo pubblicitario**, per l'**ottimizzazione della customer experience**, ma anche per la **produzione**, la **manutenzione** e perfino per la **selezione del personale**.

Sono utilizzanti nell'ambito della **riduzione del rischio nelle analisi finanziarie**, ma anche nella **riduzione del rischio** legato all'**asset management** per l'**analisi delle frodi**, e per la s**upply chain**, dove gli analytics intervengono nel **processo di la manutenzione preventiva**.
In futuro, l'utilizzo dei Big Data è previsto nella **Pubblica Amministrazione** con la creazione di un'unica struttura centralizzata di Big Data per evitare duplicazione o perdita significativa di informazioni e anche in **campo medico** per l'analisi dei dati di miliardi di individui con sistemi superveloci, ai fini dello studio delle sequenze di DNA, del genoma e delle malattie ad essi correlate e per l'efficacia dei farmaci simulando gli effetti collaterali.

Infine, l'**ambito industriale** potrà beneficiare dei Big Data per consentire la collezione e studio di dati sui clienti e realizzare prodotti e servizi ad hoc.

L'**integrazione fra Big Data e AI** è già realtà e si è concretizzata nell'ambito del **marketing** in cui le capacità di **Machine Learning** unite all'accesso a grandi dataset hanno permesso un maggiore **capacità di profilazione e di personalizzazione** nelle offerte di prodotti e servizi.

In **campo medico** l'integrazione delle due tecnologie ha affinato e ottimizzato il confronto fra cartelle cliniche consentendo deduzioni e correlazioni più approfondite fra patologie e mezzi di cura.

2. ALGORITMI E TECNICHE DI INTELLIGENZA ARTIFICIALE PER LA SICUREZZA INFORMATICA

L'Adozione delle AI permette la gestione di grandi volumi di dati analizzabili automaticamente, con la riduzione dei tempi di individuazione di anomalie e l'attivazione rapida allarmi e warning o dei processi di risposta. Applicando algoritmi di Machine Learning e i Sistemi Esperti al campo della Cybersecurity è possibile automatizzare alcune delle capacità difensive, risolvendo gli annosi problemi di: aumento della superficie di attacco e dei sistemi da proteggere, velocità della detection di un attacco, tenuta sotto controllo di molteplici parametri di security in parallelo, costi legati al personale su attività noiose e ripetitive e time-consuming, carenza di preparazione su alcuni ambiti della security.

In particolare, i **sistemi di apprendimento automatizzato** (**Machine Learning**, n.d.r.) possono essere specializzati per implementare sistemi di **rilevazione di situazioni anomale** potenzialmente pericolose (**Anomaly Detection Systems**, n.d.r.) o per identificare set e pattern di informazioni che evidenziano **comportamenti pericolosi** in atto su reti, sistemi, dati (**Behavioural Systems**, n.d.r.). Adottati a livello di Intrusion Detection System (IDS) possono costituire un **Sistema di immunità artificiale** (**Artificial Immune System**) per la rete ed i nodi costituenti, capace di individuare immediatamente una compromissione e dare l'allarme. Alternativamente i sistemi intelligenti possono essere usati per la validazione del processo di login mediante adozione di due password e un sistema di controlli splittato in due punti di controllo, per rendere impossibile il craking della procedura da parte di un attaccante (fonte: paper scientifico "Cyber Defense Using Artificial Intelligence"). Sistemi di questo tipo

possono essere usati per collaborare con esperti umani, arrivando a individuare l'85% degli attacchi e a ridurre fino a un quinto il numero di falsi positivi (segnalazioni di attacco che si rivelano innocue, n.d.r.), (fonte: Agenda Digitale, "Intelligenza artificiale per la sicurezza informatica: gli ambiti di utilizzo e i vantaggi"). La tenuta sotto controllo di reti e sistemi diventa quindi molto più efficace ed efficiente rispetto ai mezzi tradizionali potendo contare su automazione, parallelizzazione, aggiornamento continuo rispetto alle nuove minacce e su una attività continuativa h24/7 senza alcuna sosta.

Un'altra area di implementazione della AI per la cyber security è quella dei **Sistemi Esperti** usati per implementare **Sistemi di Supporto alle Decisioni** (Decision Support System - DSS). Si tratta di soluzioni composte da una base di conoscenza e da un motore inferenziale, che elabora risposte in funzione delle informazioni già note e delle situazioni contingenti agli eventi raccolti dal sistema. La velocità di detection per incidenti, la standardizzazione di approccio, l'attivazione della risposta di mitigazione remediation e recovery senza ritardi e il miglioramento continuo di ciascuna di queste fasi eleva la capacità di difesa, permettendo agli operatori di security di dedicarsi ad attività di più alto profilo, come la complessa analisi di attacchi di tipo Advanced Persistent Threat (APT). Infine, sono in corso di adozione anche i **sistemi intelligenti** ovvero componenti software capaci di la comprensione di un linguaggio di comunicazione fra agenti (Agent Communication Language - ACL), reattività (capacità di prendere alcune decisioni e di agire autonomamente) e di proattività, mediante pianificazione, riflessione e mobilità (fonte: Agenda Digitale, "Intelligenza artificiale per la sicurezza informatica: gli ambiti di utilizzo e i vantaggi"). Più agenti intelligenti distribuiti all'interno di una rete o di un ambiente circoscritto possono comunicare tra loro, correlare informazioni raccolte e addirittura attuare insieme una strategia di difesa con minima o nulla supervisione da parte di specialisti umani.

Naturalmente vi sono delle **potenziali vulnerabilità per le AI**: i set di dati per la fase di apprendimento devono essere scelti in modo appropriato, pena l'incapacità ed inadeguatezza della deduzione, ma anche lo sfruttamento delle stesse tecnologie da parte di attaccanti per "confondere" la AI che difende, secondo il paradigma di combattimento **macchina-contro-macchina** (fonte: World Economic Forum, "3 ways AI will change the nature of cyber attacks").

Inoltre, i sistemi di AI in quanto digitali, possono essere essi stessi target di attacco con il rischio di una compromissione totale o di un uso malevolo. Per questo motivo è sempre necessaria la supervisione umana sui risultati ottenuti, l'eventuale conferma delle azioni complesse da implementare e l'introduzione progressiva e continua di attività correttive volte a migliorare la resilienza complessiva a lungo termine.

3.BIG DATA ANALYTICS PER LA CYBERSECURITY

"**L'analisi dei big data** (**Big Data Analytics**, n.d.r.) si riferisce al processo di analisi o valutazione di grandi e vari volumi di dati che spesso non viene sfruttato dai normali programmi di analisi. I dati possono essere non strutturati o semi-strutturati e, in alcuni casi, potrebbero essere una combinazione di entrambi" (fonte: ISACA, "How Big Data Aids Cybersecurity").

L'**Analisi dei dati storici** consente di ottenere controlli migliori sulle minacce della sicurezza. Operativamente, l'analisi dei dati storici permette la modellazione di comportamenti e l'identificazione di pattern ricorrenti che possono costituire modelli standard e regolari dell'ambiente informatico studiato. Qualsiasi evento registrato che rappresenti una deviazione da tali standard può essere un **segnale di minaccia**.
La modellazione consente di **studiare i dati storici degli attacchi** per trovare **comportamenti ricorrenti** degli attaccanti e **prevenirne** quindi loro **future incursioni**. Questo avviene per esempio, quando l'analisi di grandi porzioni di dati di log (**Big Data log analytics,** n.d.r.) è combinata con l'**analisi JIT** (**Just in Time**) e si raccolgono informazioni sulle macchine che hanno una connessione aperta con luoghi esterni alla rete locale (fonte: research paper How Big Data Can Improve Cyber Security). Oppure l'**uso degli analytics in accoppiata a un Event Viewer** può evidenziare attività in corso di tipo sospetto e l'individuazione preventiva di anomalie e tentativi di attacco come gli iniziali cali prestazionali in negli attacchi Distributed Denial of Service DDOS. Affinché gli analytics costituiscano una conoscenza strategica è necessario "optare per **approcci** cosiddetti **'descrittivi', 'predittivi', 'prescrittivi'**, ossia sfruttando applicazioni di big data analytics attraverso le quali generare 'insights', conoscenza utile ai processi decisionali" (fonte: ZeroUno, "Cos'è big data analytics? Ecco tutto ciò che serve sapere sull'analisi dei dati").

Nella raccolta e collezione dei dati è importante eliminare quelli non essenziali che costituiscono "rumore" e interferiscono con la performance di analisi.

4.DIFESA INTELLIGENT DRIVEN È BASATA SULL'ESPERIENZA

La difesa dai cyberattacchi che fa **uso dei sistemi intelligenti coniugati** all'uso dei **Big Data** è caratterizzata da **capacità avanzate di Situational Awareness** (sistemi capaci di avere consapevolezza della situazione contingente all'ambiente digitale in esame, n.d.r.) **e da una base di conoscenza storica**, assimilabile all'esperienza specializzata, ottenuta dalla collezione e raccolta di grandi set di dati storici e dalla analisi degli stessi.

Combinando l'analisi dei Big Data e l'Apprendimento Automatico (Machine Learning) è possibile ottimizzare i risultati ottenuti singolarmente dal Machine Learning, dal Deep Learning (fonte: Deep Learning) e dalla Big Data Analytics, eseguendo analisi approfondite di dati passati e di quelli esistenti e identificando ciò che è "storicamente normale" e distinguendolo dall'inizio di un attacco.

L'**apprendimento automatico** rafforza i parametri di sicurezza informatica, perché **avvisa a ogni deviazione nella normale sequenza di eventi e attua un contrasto immediato** con azioni di risposta. In mancanza di dati storici sulla propria realtà, una organizzazione potrebbe utilizzare i dati storici di aziende dello stesso settore e prevedere scenari di attacco plausibili a cui farsi trovare preparati.

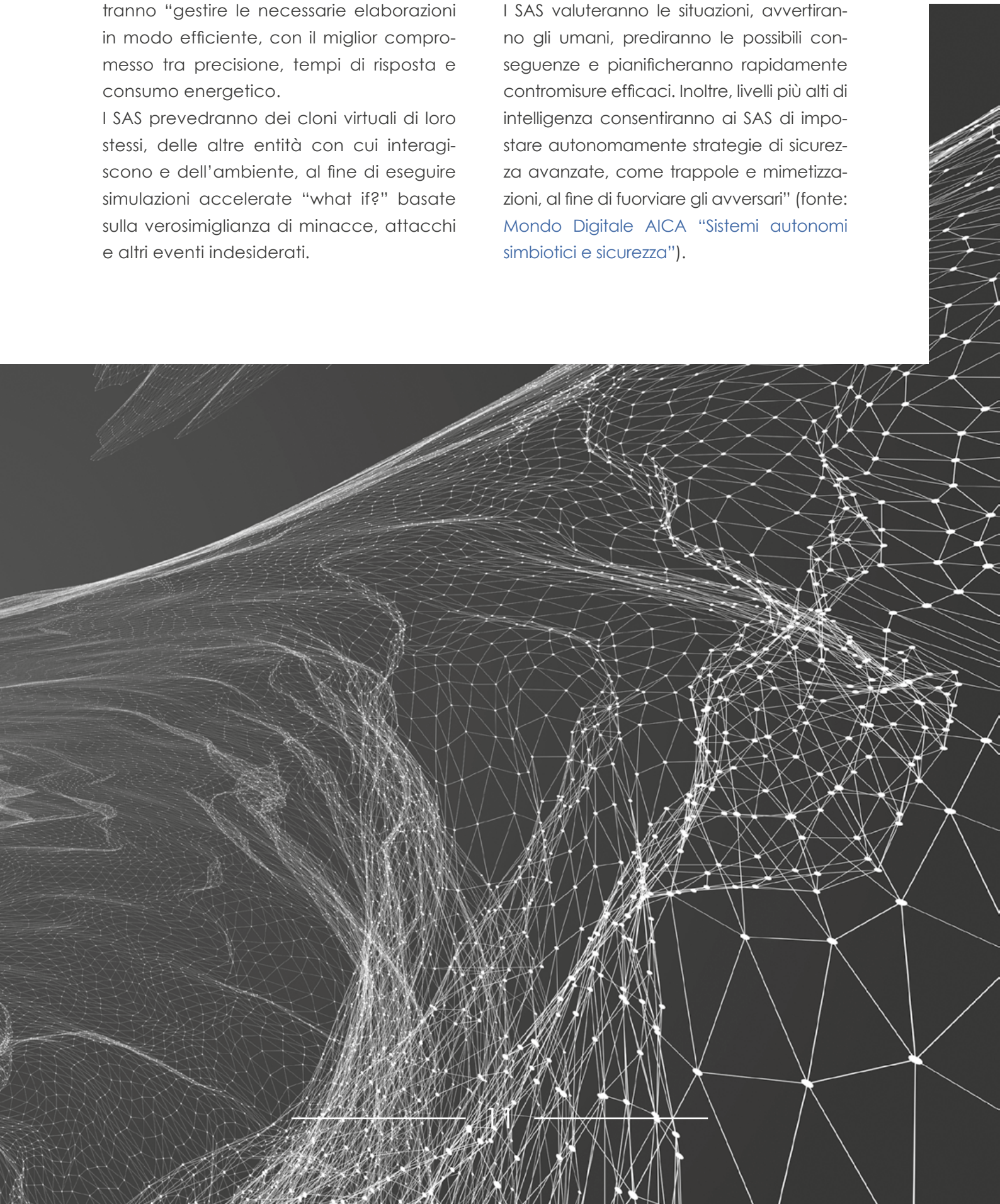
Le piattaforme **Intelligent Driven with Knowledge** possono essere **utilizzate per ogni ambito aziendale che abbia conseguenze sulla sicurezza informatica**, in modo che correli tutti gli avvenimenti, apparentemente superflui ma che invece potrebbero rientrare in uno schema di attacco. I controlli distribuiti potrebbero riguardare i sistemi Cyber Fisici costituiti dall'integrazione fra i sistemi informatici e i Sistemi di Controllo Industriale (Industrial Control Systems ICS), i processi aziendali, le attività del personale che si discostano dal rispetto di policy e procedure, gli eventi anomali relativi alla supply chain e al campo finanziario al fine di raggiungere una resillienza a 360 gradi.

Questi complessi sistemi di difesa potranno implementare in futuro i **Symbiotic Autonomous Systems**, (**SAS**), sistemi in cui la simbiosi fra uomo e macchina è più stretta della relazione uomo-strumento, i quali a loro volta potranno implementare anche capacità di simulazione per i cosiddetti scenari "what if" che analizzano le possibili evoluzioni di ogni parametro e dato che ricevono in ingresso.

Proprio i SAS, infatti, dotati di appropriata capacità di calcolo mediante combinazioni di edge, fog e cloud computing po-

tranno “gestire le necessarie elaborazioni in modo efficiente, con il miglior compromesso tra precisione, tempi di risposta e consumo energetico.
I SAS prevedranno dei cloni virtuali di loro stessi, delle altre entità con cui interagiscono e dell’ambiente, al fine di eseguire simulazioni accelerate “what if?” basate sulla verosimiglianza di minacce, attacchi e altri eventi indesiderati.

I SAS valuteranno le situazioni, avvertiranno gli umani, prediranno le possibili conseguenze e pianificheranno rapidamente contromisure efficaci. Inoltre, livelli più alti di intelligenza consentiranno ai SAS di impostare autonomamente strategie di sicurezza avanzate, come trappole e mimetizzazioni, al fine di fuorviare gli avversari” (fonte: Mondo Digitale AICA “Sistemi autonomi simbiotici e sicurezza”).



.CONCLUSIONE

Garantire la sicurezza informatica dei sistemi digitali e della loro progressione verso sistemi distribuiti complessi digitali è oggi una sfida impegnativa che può costituire una opportunità se e solo se sarà affrontata con competenza, unendo le forze della ricerca privata e accademica e le risorse umane ed economiche nell'ambito di un quadro e piano strategico nazionale. La **cyber security** è un **ambito multidisciplinare**, che richiede una solida base in **ingegneria** ed **informatica**, ma anche a livello **etico, sociale e legale**, e che già oggi presenta impatti su molti aspetti dell'esistenza umana.

La specializzazione ed adozione delle **AI** in integrazione ai **Big Data** nel campo della **sicurezza informatica** costituisce solo l'inizio di questo affascinante viaggio.

CYBEROO

CONTATTI

CYBEROO S.P.A.

Via Brigata Reggio, 37
42124 (RE)

TEL 0522.385011

WWW.CYBEROO.COM