

Indice della Tesi

Capitolo 1 Introduzione	2
1.1 Sicurezza informatica: un panorama globale.....	2
1.2 Le intelligenze artificiali come minacce emergenti.....	2
1.3 Evoluzione degli attacchi informatici basati sulle IA.....	3
Capitolo 2 Cosa sono le IA e qual'è il loro scopo.....	4
2.1 Concetti fondamentali.....	4
2.2 Differenza tra Deep Learning e Machine Learning.....	6
2.3 Descrizione delle varie tipologie di IA.....	9
2.4 Panoramica delle principali tecniche per l'apprendimento.....	
automatico delle IA.....	11
Capitolo 3 Sicurezza delle IA.....	13
3.1 Concetti di base di sicurezza informatica e le principali tecniche di difesa.....	13
3.2 Concetti di sicurezza delle IA, i rischi e le conseguenze relative ai loro attacchi.....	19
Capitolo 4 Attacchi basati sulle IA.....	22
4.1 Potenzialità a scopo offensivo.....	22
4.2 Analisi degli attacchi sulle IA già esistenti.....	25
4.3 Studio di casi e scenari in cui le IA è stata utilizzata per attacchi informatici.....	28

Capitolo 5 Contromisure e difese.....	32
5.1 Presentazione di contromisure relative ai attacchi e strategie di difesa legate ad esso.....	32
5.2 Esplorazione di tecniche innovative per difendersi utilizzando le IA stesse.....	35
5.3 Tecniche di rilevamento alle intrusioni da parte delle IA.....	38
Conclusioni e ringraziamenti.....	41
Bibliografia.....	42

Capitolo Primo

Introduzione

SOMMARIO: 1.1 Sicurezza informatica: Un panorama Globale. - 1.2 Le intelligenze artificiali come minacce emergenti. - 1.3 Evoluzione degli attacchi informatici basati sulle IA.

1.1 Sicurezza informatica: Un Panorama Globale

La sicurezza informatica, al giorno d'oggi, ha subito un'evoluzione incredibile e ha sviluppato un'importanza fondamentale nella nostra era digitale. La sua importanza si estende ben oltre la protezione di dati sensibili e privacy in generale, ma coinvolge ormai tutti i settori esistenti nell'economia moderna, nella stabilità sociale e nella fiducia nell'utilizzo delle tecnologie digitali. Con il continuo sviluppo e la continua dipendenza da reti e sistemi informatici, la protezione dei dati sensibili è diventata una priorità strategica importantissima. Le minacce, che una volta erano principalmente situate e isolate in pochi settori, adesso si sono trasformate in un panorama globale di vitale importanza per tutti. Ormai la sicurezza informatica non riguarda solo più le imprese principali del mercato ma ogni individuo che interagisce con il mondo digitale.

1.2 Le intelligenze artificiali come minacce emergenti

Con l'arrivo delle intelligenze artificiali (IA), nel mondo della sicurezza informatica bisogna iniziare ad affrontare nuovi problemi di ogni genere. Le IA, che inizialmente furono sviluppate per migliorare la nostra vita

quotidiana, iniziarono poi in seguito ad essere utilizzate in maniera malevola, introducendo nel mondo della sicurezza informatica nuovi tipi di minacce. L'ingegnosità di questi attacchi si basa soprattutto sull'utilizzo degli algoritmi per l'apprendimento automatico e di reti neurali per creare attacchi evoluti che possono eludere e danneggiare le difese tradizionali. Questo nuovo tipo di sviluppo richiede una revisione approfondita di quelle che sono le strategie di sicurezza informatica per affrontare minacce emergenti sempre più intelligenti e adattative.

1.3 Evoluzione degli attacchi informatici basati sulle IA

L'evoluzione degli attacchi informatici, alimentati dalle IA, è in continua crescita e non smette mai di progredire ogni giorno che passa.

L'uso di tecniche come il Machine Learning ha introdotto una grandissima flessibilità, mai viste in precedenza, nell'attaccare gli obiettivi selezionati in maniera efficace, consentendo ai aggressori di adattare in maniera incredibile le loro strategie rendendo le contromisure inefficaci. Questo tipo di sviluppo richiede un costante adeguamento delle strategie utilizzate per difendersi in modo che possano anticipare e rispondere a tutti i tipi di minacce emergenti già esistenti e a quelle che arriveranno. In sintesi, la sicurezza informatica è diventata una priorità globale in tutti i tipi di contesto, soprattutto dalle minacce dell'arrivo delle IA e la comprensione di questi aspetti chiave è essenziale per lo sviluppo di nuovi approcci di sicurezza avanzati con il fine di proteggere la nostra società digitale in modo molto efficace.

Capitolo Secondo

Cosa sono le IA e qual'è il loro scopo

SOMMARIO: 2.1 Concetti fondamentali. – 2.2 Differenza tra Deep Learning e Machine Learning. – 2.3 Descrizione delle varie tipologie di IA – 2.4 Panoramica delle principali tecniche per l'apprendimento automatico delle IA.

2.1 Concetti fondamentali

Le Intelligenze Artificiali (IA) nel mondo odierno rappresentano il picco massimo dell'innovazione tecnologica, una frontiera che affronta e attraversa la convergenza dei algoritmi più avanzati e significativi che ci sono nell'odierno mondo digitale. Oltre a questi attraversano nelle loro specifiche anche altri settori come la potenza di calcolo e la possibilità di avere accesso a un enorme quantità di dati.

Definire le IA va oltre l'etichettatura di macchine intelligenti, infatti sono molto più di questo; si tratta di una vera e propria immersione profonda in un mondo in cui le macchine imparano, apprendono, ragionano e si adattano a tutto, sfidando i confini dell'intelligenza umana.

“L'intelligenza artificiale è la scienza che si occupa di far fare alle macchine, le cose che richiederebbero l'intelligenza se fossero fatte dagli esseri umani.”¹

Uno degli scopi che principalmente hanno le IA è migliorare l'efficienza operativa e produttiva attraverso l'automazione di compiti che sarebbero incredibilmente ripetitivi e molto dispendiosi in termini di tempo.

L'automazione, in questo caso, consente alle IA di eseguire quelle attività produttive e non che richiederebbero una grandissima precisione e velocità, liberando risorse umane per compiti più creativi e strategici.

Le IA mirano a fornire anche un vantaggio predittivo attraverso l'analisi avanzata dei dati. Attraverso algoritmi sofisticati, le IA sono in grado di anticipare modelli, tendenze e altre attività di sviluppo consentendo

¹ Marvin Minsky

all'operatore di prendere decisioni basate su dati accurati e molto dettagliati con grande precisione.

Un aspetto cruciale delle IA è soprattutto la loro capacità di adattarsi in tempo reale alle dinamiche di qualsiasi ambiente vengano utilizzate. Infatti, gli algoritmi di apprendimento automatico consentono alle macchine di regolare e addirittura modificare il proprio comportamento in risposta a nuove informazioni che sono state inserite a riguardo, migliorando costantemente e in maniera importante le prestazioni.

In un contesto industriale, le IA giocano un ruolo chiave e molto importante nella trasformazione dei processi produttivi. Dall'ottimizzazione delle catene di approvvigionamento fino ad arrivare alla manutenzione predittiva delle macchine, le IA infatti stanno rivoluzionando ogni giorno che passa il modo in cui le aziende, in qualsiasi settore, affrontano le loro sfide quotidiane iniziando ad essere quasi indispensabili.

Settori come la medicina beneficiano già da tempo delle IA per l'analisi di immagini mediche, la diagnosi di malattie e addirittura vengono utilizzate per la personalizzazione dei trattamenti. L'IA viene infatti utilizzata per assistere i professionisti medici nella presa di decisioni complesse, migliorando l'efficacia della ricerca e delle cure stesse.

Il futuro delle IA infatti promette grandi miglioramenti con l'integrazione di tecniche sempre più difficili e complesse come l'apprendimento profondo, la comprensione del linguaggio umano (naturale) e la capacità di ragionamento simbolico. L'evoluzione delle IA continua a modificare il nostro modo di vivere, lavorare e addirittura interagire con tutto il mondo digitale.

Nonostante tutti benefici le IA comunque sollevano comunque grandi questioni etiche e sociali molto significative. La responsabilità nella progettazione di algoritmi, la trasparenza delle decisioni e la possibilità a tutti di poter accedere alle tecnologie emergenti diventano aspetti importantissimi nel dibattito sull'adozione e lo sviluppo delle IA con temi tutt'oggi caldi e pieni di discussioni.

2.2 Differenza tra Deep Learning e Machine Learning.

Due pilastri fondamentali che costituiscono questo studio sulle IA sono il Machine Learning (ML) e il Deep Learning (DL), entrambi con caratteristiche uniche che li distinguono tra loro nel vasto panorama delle IA.

Il Machine Learning esiste da diversi anni ma solo ultimamente sta diventando di fondamentale importanza, questo grazie alle evoluzioni delle nuove tecnologie e dei computer sempre più capaci di elaborare velocemente una grandissima quantità di dati.

Queste frasi qua sotto, sono due definizioni date da dei famosi personaggi legati a questo ambito. Iniziamo con quella proposta da Arthur Samuel, un ricercatore informatico leader e pioniere nell'ambito dell'intelligenza artificiale che ha descritto il ML come *“ il settore di studi che fornisce al computer l'abilità al computer di apprendere senza essere esplicitamente programmato. ”* La seconda invece, è quella di Tom Mitchell, un informatico che ha fornito una definizione molto più moderna del ML, la descritto come *“ a computer programme is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance in tasks T, as measured by P, improves with experience E.”*²

Quindi in sostanza il ML, è un campo dell'IA che consente alle macchine di apprendere dati senza essere programmate in maniera dettagliata. Questo approccio sfrutta soprattutto svariati modelli matematici per fare previsioni attendibili e prendere decisioni basate su dati di addestramento, permettendo alle macchine di migliorare le loro prestazioni.

Principalmente le tipologie di ML si suddividono in tre modelli:

1. Supervised Learning

Modello che viene addestrato su un set di dati etichettato, dove gli input vengono associati ad output molto noti. Il modello apprende le relazioni tra essi consentendo predizioni molto accurate sui nuovi dati.

2. Unsupervised Learning

Contrariamente al modello descritto in precedenza, questo è esposto principalmente a dati non etichettati. Il suo compito è di

² G. Balinggan, 2019. [Online]. Available: [http://www.differencebetween.net/science/differencebetween-artificialintelligence-and-human-intelligence/..](http://www.differencebetween.net/science/differencebetween-artificialintelligence-and-human-intelligence/)

individuare pattern, relazioni e strutture dei dati senza avere indicazioni dall'esterno.

3. *Reinforcement Learning*

Modello basato sulla ricompensa, implica infatti che l'entità apprenda attraverso l'interazione con un ambiente specifico. Riceve ricompense o penalità in base alle azioni che commette, adattandosi poi in conseguenza alle risposte.

E' importante distinguere chiaramente tra il concetto di Deep Learning e il concetto di Machine Learning, riconoscendo che sebbene entrambi contribuiscano allo sviluppo dell'intelligenza artificiale, rappresentano due concetti chiari e distinti tra di loro che portano contributi unici e specifici al campo.

Il Deep Learning rappresenta infatti una sottocategoria del Machine Learning caratterizzata dall'impiego di reti neurali artificiali dotate di diversi strati nascosti. Questo approccio consente all'elaborazione complessa e non lineare dei dati, permettendo così di identificare elementi distintivi direttamente da fonti di dati non elaborati, quali pixel, scansioni mediche o sequenze biologiche e tanto altro...³

Queste reti infatti, sono ispirate al cervello umano, presentando multi strati di neuroni artificiali, consentendo una rappresentazione molto più ricca, elaborata ed astratta dei dati.

Approfondiamo meglio i concetti chiave del DL.

Il cuore del DL è costituito dalle reti neurali profonde, che sono modelli composti da strati multipli di neuroni interconnessi tra di loro. Queste reti infatti sono organizzate in strati di input, hidden e layers e output, consentendo una modellazione molto complessa delle relazioni nei dati.

Suddividiamo le DL in varie architetture molto popolari nel web:

1. *Reti Neurali Convoluzionali (CNN)*

Sono reti che vengono utilizzate per l'elaborazione di immagini e di video, le CNN sono infatti specializzate nella rilevazione di pattern spaziali. Le CNN vengono principalmente utilizzate in applicazione di visione artificiale e infatti stanno venendo

³ Nvidia, «Nvidia Jetson,» [Online]. Available: <https://www.nvidia.com/it/autonomous-machines/embedded-systems>.

implementate con svariati VR per il riconoscimento di oggetti e il rilevamento di caratteristiche visive.

2. *Reti Neurali Ricorrenti (RNN) e le architetture Long Short-Term Memory (LSTM)*

Queste reti invece sono molto adatte per dati sequenziali, le RNN e le LSTM vengono spesso utilizzate in ambiti come il riconoscimento del linguaggio naturale e la generazione di testi. Queste architetture sono in grado di mantenere memoria a lungo termine, gestendo in maniera significativa sequenze di dati in maniera decisamente meglio accurata delle reti neurali tradizionali.

3. *Reti Generative Avversarie (GAN)*

Le GAN vengono utilizzate per la generazione di nuovi dati realistici, esse comprendono un generatore e un discriminatore che sono allenati in maniera simultanea creando una sorta di sfida tra di loro per migliorare la qualità dei dati che vengono generati.

Nel mondo odierno il DL ha rivoluzionato sotto diversi aspetti lo studio delle IA, infatti tutt'oggi viene applicato in svariati contesti come la visione artificiale, il riconoscimento del linguaggio naturale e soprattutto in ambito medico.

In conclusione, mentre il ML viene visto come un approccio molto versatile e facilmente adattabile a una vasta gamma di situazioni, proprio il Deep Learning emerge in quanto complessità perchè richiede una modellazione molto più sofisticata.

L'armonia tra di loro è la chiave per affermarsi nei progetti futuri dove le IA inizieranno ad evolversi in maniera significativa. Molto probabilmente il futuro delle IA vedrà una crescita sinergica di ML e DL, in cui gli algoritmi si integreranno con la potenza di modellazione delle reti neurali profonde aprendo strada a un mondo in cui l'IA continua a superare i confini della tecnologia.

2.3 Descrizione delle varie tipologie di IA

Dopo aver compreso come apprendono le IA, ora vediamo come si suddividono e come vengono utilizzate in svariati ambiti.

In questo capitolo esploriamo le varie sfaccettature dell'IA, offrendo un'analisi molto profonda delle tipologie e presentando concetti chiave fondamentali estrandole da opere autorevoli nel panorama odierno delle IA.

L'IA viene creata come disciplina con l'obiettivo di emulare l'intelligenza umana, si articola in varie tipologie, ognuna con caratteristiche uniche nel suo genere. Attraverso vari approfondimenti e studi riportati in questo specifico settore possiamo vedere un panorama sempre più ricco e complesso di informazioni.

Possiamo suddividere le IA in 5 grandi categorie:

1. IA Debole e IA Forte

Nils J. Nilsson, traccia tra i primi una distinzione cruciale della differenza tra IA debole e IA forte. Secondo Nilsson proprio la Intelligenza Artificiale debole, o IA specializzata, viene progettata per compiti molto specifici senza avere una comprensione globale o una consapevolezza, mentre invece la IA forte aspira proprio a una vera e propria forma di intelligenza equiparabile a quella dell'essere umano. Questa dicotomia fornisce le fondamenta per esplorare quelle che sono le varie applicazioni dell'IA in settori in fase di sviluppo come la guida autonoma o addirittura il riconoscimento vocale.⁴

2. IA Simbolica e Connessionista

L'IA simbolica è basata su regole e rappresentazioni esplicite mentre invece l'IA connessionista adotta delle reti neurali e apprendimento implicito. Questo dibattito tra l'IA simbolica e l'IA connessionista riflette le varie filosofie che guidano lo sviluppo di algoritmi e modelli di ricerca per avere delle IA in futuro molto più sviluppate e flessibili.

3. IA Basata sulla Conoscenza

Questo modello di IA impiega la rappresentazione di conoscenza per prendere decisioni e risolvere i problemi. La capacità di

⁴ Nils J. Nilsson, "Artificial Intelligence: A New Synthesis"

modellare il mondo in termini di conoscenza dichiarativa è essenziale per affrontare situazioni molto complesse, infatti risulta di vitale importanza rendere questo paradigma funzionante per applicazioni medica come la diagnosi, le terapie e addirittura per altri settori come la guida autonoma. Pionieri di questo modello di IA sono stati Stuart Russell e Peter Norvig che esplorano in maniera molto significativa l'IA basata sulla conoscenza.⁵

4. *IA Evolutiva e Swarm Intelligence*

L'IA evolutiva prende ispirazione sul principio dell'evoluzione biologica per migliorare le soluzioni nel tempo, mentre la swarm intelligence fa affidamento al comportamento degli insetti sociali per affrontare e poi risolvere i problemi di elevata complessità in maniera collaborativa.

5. *IA Quantistica*

Scott Aaronson, per la prima volta, ci introduce in maniera significativa al concetto di IA quantistica. Questa tipologia di IA si basa sullo sfruttamento dei principi della meccanica quantistica per eseguire calcoli in maniera decisamente più veloce di un classico computer. L'IA offre prospettive decisamente rivoluzionarie per la soluzione di problemi molto complessi, come la simulazione di molecole per la progettazione di farmaci e altri progetti in ambito medico.⁶

Esplorando le diverse tipologie di IA abbiamo creato un panorama approfondito e stimolante di un percorso affascinante nel mondo delle IA. Ovviamente, queste tipologie principali di IA non si fermano a queste cinque, ma qua ho voluto riportare solo le categorie di maggiore rilievo in questo momento. Il futuro sicuramente promette ulteriore crescita in questo ambito e ci saranno sicuramente sviluppi che oggi nemmeno possiamo immaginare.

L'intelligenza artificiale è ormai un settore vasto e dinamico, alimentato da una ricca ricerca e da un continuo sviluppo in tutto il mondo.

⁵ Stuart Russell and Peter Norvig, "Artificial Intelligence: A Modern Approach"

⁶ Scott Aaronson, "Quantum Computing Since Democritus"

2.4 - Paronamica delle principali tecniche per l'apprendimento automatico delle IA.

Questo capitolo offre uno sguardo approfondito e analitico di quelle che sono le metodologie fondamentali che guidano l'evoluzione dell'apprendimento automatico delle IA.

Esploriamo ora in maniera approfondita le tecniche chiave di questo argomento, mettendo in evidenza il loro impatto anche sulle IA moderne.

Michael Negnevitsky ha suddiviso in maniera molto significativa due aspetti fondamentali dell'apprendimento automatico. Il primo è il Supervised Learning che sarebbe l'apprendimento supervisionato, cioè una tecnica in cui il modello di IA è addestrato su un insieme di dati che vengono etichettati. Il sistema impara a fare delle previsioni o impara a svolgere delle azioni basandosi su dati forniti in input dall'utente, minimizzando l'errore tra previsioni ed effettivi.

Questa tecnica viene utilizzata in svariati ambiti come il riconoscimento di immagini e anche per la comprensione da parte della macchina del linguaggio naturale. Il secondo metodo invece è l'Unsupervised Learning cioè l'apprendimento non supervisionato, che a differenza del precedente esso coinvolge l'addestramento basato su modelli non etichettati cercando pattern e strutture nascoste di dati. Un esempio di questo metodo d'apprendimento è l'analisi esplorativa dei dati e anche la scoperta di relazioni sconosciute.⁷

Tuttavia esistono anche diversi metodi di apprendimento automatico, uno di questi è il Reinforcement Learning che sarebbe l'apprendimento per rinforzo. Esso viene basato su un'idea di far apprendere in un ambiente dinamico guadagnando o perdendo in base alle azioni che vengono intraprese. L'agente che impara con questo metodo capisce quale azione deve svolgere in una determinata situazione in maniera da massimizzare il rendimento a lungo termine. Questa tecnica viene usata per applicazioni come i veicoli autonomi e anche per i giochi strategici che nell'ultimo periodo stanno spopolando sul web.

Infine abbiamo ancora due categorie di apprendimento automatico che sono il Deep Learning che abbiamo già affrontato in precedenza e per ultimo il Transfer Learning che sarebbe l'apprendimento di trasferimento,

⁷ Negnevitsky, Michael. "Artificial Intelligence Techniques." Editore XYZ, 2022. Capitolo 2.4.

ovvero un modello di apprendimento dove vengono trasferite le conoscenze apprese da un compito all'altro.

Ad esempio, un modello di questo tipo viene addestrato su un gran numero di dati di immagini che ne beneficiano per svolgere compiti specifici senza richiedere un enorme set di dati aggiuntivi. Questa tecnica diventa di fondamentale importanza quando si ha a che fare con dati di disponibilità molto limitata.

Queste metodologie sono le fondamenta su cui mettere le basi per l'intelligenza artificiale moderna. L'implementazione di queste tecniche offre un potenziale incredibile nel trasformare settori chiave e nel migliorare la vita di tutti i giorni in maniera significativa.

Tuttavia, diventa di vitale importanza mantenere una trasparenza e un'equità nel adottare in sicurezza queste tecniche di apprendimento dato che attualmente queste vengono applicate a settori di grande importanza.

Queste tecniche vengono applicate alle IA per studiare e analizzare mercati finanziari, prevenzione dei rischi le tendenze e anche settori per esempio la ricerca e sviluppo ne beneficiano dato che grazie alle IA si accelerano le scoperte di nuovi prodotti e di nuovi metodi di lavoro più produttivi.

Capitolo Terzo

Sicurezza delle IA

SOMMARIO: 3.1 Concetti di base di sicurezza informatica e le principali tecniche di difesa. – 3.2 Concetti di sicurezza delle IA e i rischi relativi ai loro attacchi. – 3.3 Varie categorie di minacce legate alle IA e le relative conseguenze.

3.1 Concetti di base di sicurezza informatica e le principali tecniche di difesa

Nel terzo capitolo, vengono affrontati i principali concetti di sicurezza informatica e loro le loro relative tecniche di difesa sia legate alle IA che non , e verranno esplorati in dettaglio come esse interagiscono col mondo odierno.

3.1.1 Concetti di base di sicurezza informatica

Nel 1988 Robert Morris, lanciò uno dei primi concetti di sicurezza informatica e lo espose come: “ *La sicurezza informatica è quella disciplina che si occupa della protezione dei sistemi informatici, hardware, software e dati da minacce esterne e interne.* ” ⁸

Nel contesto legato alle IA, questo concetto viene amplificato perchè le macchine intelligenti elaborano una quantità industriale di dati sia sensibili che non sensibili prendendo decisioni in maniera autonoma.

Come studiato nel percorso scolastico possiamo quindi racchiudere in sintesi la sicurezza informatica in tre macro concetti chiave che sono l'autenticazione, l'autorizzazione e la crittografia.

Nel primo caso l'autenticazione rappresenta proprio il processo che va a verificare l'identità di un utente, di una macchina o di un tipico ambiente di lavoro informatico che tenta di accedere a un determinato sistema.

⁸ Robert Morris

Riprendendo il contesto delle IA, l'autenticazione risulta fondamentale per verificare che solamente gli utenti che sono stati autorizzati possano accedere o interagire con i dati disponibili.

L'autenticazione può essere imposta in vari modi ma tra le principali vediamo la Single-Factor Authentication, che sarebbe la forma più comune di autenticazione che coinvolge l'uso di una password segreta come nelle classiche pagine del web o nei giochi tradizionali online.

Un altro tipo di autenticazione è quella Two-Factor Authentication ovvero la 2FA che sarebbe un'autenticazione a doppio fattore che, oltre alla classica password, l'utente deve ricevere un codice temporaneo (principalmente viene trasmesso tramite un sms o da un app di autenticazione) per poter accedere ai dati, fornendo un ulteriore livello di sicurezza.

Per ultimo, abbiamo l'autenticazione biometrica ovvero un tipo di autenticazione basato o sul riconoscimento facciale o sul finger print. Le sue principali caratteristiche sono l'uso di caratteristiche uniche per ogni individuo richiedendo un fattore di unicità per poter accedere ai dati e aggiungere un fattore di comodità non indifferente nel mondo digitale odierno. Ormai le principali app di multi-banking, posteID e altri servizi legati all'identità pubblica utilizzano questo tipo di autorizzazione rendendo molto più facile e veloce l'accesso ai servizi disponibili.

Dopo aver affrontato questi tre aspetti dell'autenticazione ci poniamo una domanda, ma perchè non viene allora utilizzato sempre il modello di riconoscimento biometrico?

Semplice, il fatto che nei tre modelli elencati in precedenza c'è un fattore legato al costo dello sviluppo del software e alla sicurezza non indifferente, infatti nel primo caso un semplice accesso con password richiederà sicuramente un costo molto inferiore rispetto a un 2FA, mentre invece uno sviluppo con riconoscimento biometrico ha un costo notevolmente superiore a questi due.

Dopo aver affrontato tutti questi concetti di autenticazione ora parliamo del fattore che determina i diritti e i privilegi di accesso dell'utente che viene autenticato, ovvero l'autorizzazione.

L'autorizzazione, soprattutto nei concetti legati alle IA è cruciale per limitarne l'accesso e le funzionalità e risulta fondamentale per proteggere i nostri dati ed evitare manipolazioni non autorizzate.

Tra i principali tipi di autorizzazione troviamo il modello basato su ruoli (Role-Based Access Control ovvero RBAC) che assegna i diritti di accesso in base al ruolo che ha l'utente quando viene riconosciuto. Un classico esempio è quello di un'azienda dove l'amministratore avrà sicuramente privilegi diversi da quello di un utente normale o di un classico collaboratore.

Un altro tipo importante di autorizzazione è quello basato su attributi (Attribute-Based Access Control che sarebbe il classico ABAC). In questo specifico modello di autorizzazione vengono messi attributi specifici all'utente per determinarne l'accesso. Ad esempio, un utente che vuole loggare può ottenere l'accesso solo se si trova in una posizione geografica specifica che viene preimpostata quando viene implementato il software.⁹

Esistono tanti altri modelli di autorizzazione e autenticazione ma in questa tesi specifica ho voluto sviluppare solo quelli che ritengo più importanti e quelli che ritenevo davvero i più necessari per continuare il percorso legato alle IA.

Ora affrontiamo il concetto principale della crittografia, ovvero un pilastro fondamentale nella sicurezza informatica.

La crittografia rimane uno dei mezzi sicuri a cui affidarci per proteggere la riservatezza dei dati durante la loro trasmissione. In un contesto moderno dove le IA elaborano i dati e operano anche molto spesso su dati sensibili, la crittografia diventa una barriera necessaria contro le potenziali minacce e accessi non autorizzati. Affrontiamo nel dettaglio i concetti chiave della crittografia come abbiamo appreso nel nostro percorso scolastico:

1. Crittografia a Chiave Simmetrica

Questo tipo di crittografia coinvolge l'uso di una singola chiave per cifrare e decifrare dati. Sicuramente è tra i più rapidi ed efficaci ma il principale dilemma di questo utilizzo è sicuramente la distribuzione sicura delle chiavi. Vengono utilizzati dei veri e propri algoritmi di decodifica come abbiamo studiato come il DES, IDEA, 3DES e altri ancora... Questi algoritmi si basano sull'uso di una chiave privata unica per entrambi i processi di codifica e decodifica, oppure utilizzano chiavi private differenti che tuttavia sono strettamente connesse tra loro con un vero e proprio processo di derivazione.

⁹ <https://cips.it/autenticazione-vs-autorizzazione-come-comprendere-la-differenza/>

2. Crittografia a Chiave Pubblica / Privata

Questo approccio utilizza una coppia di chiavi una pubblica e una privata, dove quella pubblica viene utilizzata principalmente per cifrare i dati e poi viene usata una chiave privata per decifrarli. La chiave pubblica infatti può essere condivisa liberamente mentre invece quella privata deve rimanere assolutamente segreta dato che viene utilizzata per decifrare i dati. Questo metodo viene spesso utilizzato per garantire la sicurezza nelle comunicazioni, soprattutto quelle online, offrendo alle aziende e ai privati con cui comunicano una soluzione più elaborata rispetto alla crittografia simmetrica. La crittografia a chiave pubblica viene spesso utilizzata per il traffico di posta elettronica, ad esempio nello standard crittografico S/MIME ma anche nei protocolli crittografici come SSH, SSL/TLS e HTTPS.

3. Crittografia End-to-End

La crittografia end-to-end, spesso anche abbreviata e nota come “etee”, garantisce agli utenti coinvolti nelle comunicazioni che i loro dati rimangano protetti da accessi non autorizzati, anche in caso di attacchi ai server di servizi che utilizziamo tutti i giorni come Whatsapp, Telegram ecc... Questo metodo consente di cifrare i messaggi in modo che solo i destinatari previsti possano leggerli, rendendo i server aziendali meri trasmettitori incapaci di interpretare i contenuti scambiati all'interno. In sintesi, la crittografia etee cifra i dati dall'origine al destinatario finale, prevedendo l'accesso da parte di terzi.

4. Firma Digitale

La firma digitale rappresenta un'importante funzionalità offerta dalla crittografia a chiave pubblica, essenziale per confermare sia l'autenticità di documenti o messaggi sia l'identità di chi li invia. Tale processo prevede l'uso di una firma creata dal mittente e la successiva verifica attraverso una chiave pubblica apposita. La generazione di una firma digitale richiede il ricorso a enti di certificazione riconosciuti e supervisionati dall'AGID, che forniscono i necessari certificati digitali e di autenticazione.

5. Crittografia Quantistica

Considerando l'evoluzione delle IA e aspetti futuri, la crittografia quantistica potrebbe diventare sempre più rilevante dato che si basa su principi della fisica quantistica dove insieme offrono una sicurezza basata su leggi fisiche fondamentali. In sostanza il meccanismo della crittografia quantistica si lega a due utenti che dentro un canale di comunicazione comune possono generare stringhe di informazioni condivise e segrete, utilizzabili come chiavi segrete per la crittografia standard, al fine di garantire comunicazioni sicure. La proprietà più singolare non è solo la capacità di proteggere le comunicazioni, ma anche quella di permettere il rilevamento di qualche terza parte che cercasse di penetrare nella trasmissione e acquisire la conoscenza della chiave segreta.¹⁰

In conclusione, la crittografia rimane un pilastro fondamentale nella creazione di un ambiente sicuro, garantendo che i dati e i modelli utilizzati rimangano inalterati e diventino inaccessibili da utenti non autorizzati che tentano operazioni di manipolazioni malevoli.

La sua evoluzione può infatti diventare un fattore chiave nel progresso che stanno facendo le IA in questi ultimi anni.

¹⁰ <https://www.cybersecurity360.it/soluzioni-aziendali/crittografia-quantistica-cose-e-come-usarla-per-garantire-massima-protezione-ai-dati-sensibili/>

3.1.2 Principali tecniche di difesa

Le IA hanno rivoluzionato tantissimi settori, uno in particolare è come difendersi da attacchi informatici. I attacchi informatici crescono di giorno in giorno aumentando la loro diffusione e la loro complessità, quindi emergono esigenze legate a come difendersi da questi attacchi.

Nel contesto della sicurezza informatica, l'adozione di un approccio stratificato e dettagliato è fondamentale per proteggere efficacemente le risorse digitali dalle minacce sempre più sofisticate. Un esempio specifico di questa strategia è l'implementazione di firewall di nuova generazione (NFW), che differiscono dai tradizionali firewall per la loro capacità di filtrare il traffico di rete non solo in base agli indirizzi IP e alle porte, ma anche analizzando il contenuto dei pacchetti dati per identificare e bloccare attività malevole basate su applicazioni specifiche.

Parallelamente, l'integrazione di sistemi di prevenzione delle intrusioni (IPS) che utilizzano tecniche avanzate di rilevamento basate su firme, anomalie e comportamenti sospetti, consente di identificare attivamente tentativi di intrusione in tempo reale, bloccandoli prima che causino danni.

Questi sistemi sono spesso arricchiti con intelligenza artificiale per migliorare l'accuratezza del rilevamento delle minacce.

Il sandboxing emerge come un'altra tattica specifica, offrendo un ambiente controllato dove i file sospetti possono essere eseguiti senza rischio per la rete principale. Questo permette un'analisi dettagliata del comportamento del file, identificando malware che potrebbe non essere rilevato attraverso metodi convenzionali.

L'autenticazione a più fattori (MFA) rappresenta una misura di sicurezza critica, andando oltre le semplici password per richiedere agli utenti di verificare la loro identità attraverso più metodi prima di concedere l'accesso ai sistemi. L'uso di token hardware, autenticazione biometrica o codici inviati via SMS sono esempi di fattori aggiuntivi che aumentano significativamente la sicurezza degli accessi.

Proseguendo su questa strada, la formazione specifica dei dipendenti su minacce come il phishing e le tecniche di ingegneria sociale è vitale. Programmi di formazione personalizzati che includono simulazioni di attacchi phishing e workshop interattivi possono aumentare la consapevolezza dei dipendenti trasformandoli da potenziali vettori di attacco a primi difensori contro le minacce informatiche.

In seguito, dobbiamo sempre usare un antivirus e mantenerlo sempre aggiornato. Ne esistono di vari tipi, tra questi vediamo principalmente gli EDR (Endpoint Detection and Response) che costituiscono le versioni più avanzate, infatti sono in grado di analizzare attivamente e continuamente quanto accade nei nostri pc, rilevando eventuali anomalie.

Invece per chi tratta grandi quantità di dati o chi fornisce servizi web, può essere utile appoggiarsi ad alcuni servizi cloud e usare strumenti per l'analisi del traffico in entrata ed uscita della propria rete (firewall).

Anche le versioni dei database SQL dovranno sempre essere aggiornate e monitorate per evitare che qualcuno faccia breccia nel sistema.

Un altro metodo è sicuramente il prestare attenzione ai link che riceviamo, sia per messaggio che per e-mail, ed evitare di connettersi a reti "aperte", ovvero principalmente pubbliche, e soprattutto aumentare l'uso di sicurezza e autenticazione a due fattori o biometrica per garantire una solidità dei dati che trattiamo.

Questi principi sono alla base della sicurezza informatica.

La maggior parte dei siti comunque ha già un livello di crittografia già avanzato (HTTPS dove la S sta per Secure) ma ci possono essere anche siti non sicuri, senza la S dove in quel caso se accedessimo a un sito HTTP da un wi-fi pubblico possiamo rischiare tantissimo, infatti le informazioni che scambiamo potrebbero essere rubate dato che non sono crittografate (carta d'identità, carte bancarie ecc...).

Il metodo più sicuro per proteggerci in questo caso specifico è l'utilizzo di una VPN, dato che serve proprio a crittografare tutti i flussi di dati. Oltretutto la VPN applica un secondo livello di sicurezza che possiamo immaginare come un tunnel criptato che avrà un nuovo percorso tra noi e il sito, infatti oltre al file criptato, avrà un ulteriore livello di sicurezza e come se non bastasse in aggiunta a questi tunnel creati dalla VPN creerà la nostra connessione anonima (infatti possiamo collegarci con un IP di un altro paese).

Per quanto siano disponibili vari strumenti e metodi di difesa, ricordiamo che principalmente non esistono metodi sicuri al cento per cento contro gli attacchi.

Come sempre la consapevolezza e il buon senso devono essere la guida dei nostri comportamenti, soprattutto in rete dato che nel web si combattono ogni giorno lotte tra attaccanti e difensori e quello che possiamo fare e cercare di agire con moltissima attenzione cercando di essere pronti a far fronte ad ogni conseguenza.¹¹

3.2 Concetti di sicurezza delle IA , i rischi e le conseguenze relative ai loro attacchi

Nel sempre più mutevole panorama tecnologico, la crescita di strumenti di difesa basati sulle IA ha portato opportunità senza precedenti. Tuttavia, questo notevole avanzamento della tecnologia non è giunto senza la sua serie di sfide, soprattutto legate alla sicurezza informatica. La protezione dei sistemi basati sulle IA non è solo una questione di salvaguardia dell'integrità dei dati ma anche una continua sfida contro le minacce emergenti che potrebbero compromettere la nostra sicurezza.

Nel concetto di sicurezza delle IA si considerano principalmente i tre grandi pilastri della sicurezza informatica che sono l'integrità, la disponibilità e la confidenzialità. Nel primo caso l'integrità si basa sulla correttezza e l'affidabilità delle decisioni che l'IA prende dopo aver ricevuto in input le istruzioni necessarie e si basa sulla lavorazione di dati che sono sicuramente accertati e che non sono stati compromessi.

Invece, la disponibilità assicura che i sistemi delle IA siano sempre operative, aggiornate e che siano resistenti a qualsiasi tipo di sabotaggio. Infine, la confidenzialità riguarda la difesa dei dati sensibili che le IA hanno e che elaborano continuamente, specialmente di quei dati che sono legati all'azienda in senso stretto e privato o a quei dati personali che non devono essere divulgati.

¹¹ <https://www.geopop.it/cosa-sono-gli-attacchi-informatici-e-quali-sono-le-principali-tipologie-e-le-piu-diffuse/>

Un aspetto chiave della sicurezza informatica delle IA è che i modelli che usano vengono addestrati in modo sicuro. Diventa di vitale importanza evitare che i dati vengano manipolati per evitare che ci siano comportamenti errati o non sicuri nei modelli di addestramento. Questo include, ad esempio, l'avvelenamento dei dati dove gli attaccanti inseriscono in maniera malevola delle informazioni false nel dataset di addestramento per portare l'IA a rispondere in maniera errata o in maniera irregolare.

Risulta cruciale condurre analisi corrette e regolari delle vulnerabilità nei modelli IA proprio per identificare queste azioni e apportare correzioni alle loro debolezze prima che possano essere utilizzate da qualche utente esterno a scopo di danneggiarle o disturbarle.

Sempre più aziende ormai usano le IA e quindi esse vengono addestrate proprio su dati aziendali e per tanto devono essere le aziende stesse a essere consapevoli che tali dati potrebbero essere esposti a rischi. Start-Up come Protect AI sta affrontando il problema direttamente su queste dinamiche già esistenti, cercando di fornire soluzioni che aiutino le aziende stesse a tracciare le componenti dei loro sistemi IA attraverso un "Machine-Learning bill of materials". Questa piattaforma non solo identifica potenziali violazioni della sicurezza, ma anche attacchi con iniezione di codice maligno.

Il panorama moderno infatti vede come nuova minaccia emergente la "Injection Prompt" dove gli hacker manipolano i sistemi di intelligenza artificiale per divulgare le informazioni sensibili e venderle al miglior offerente.¹²

Gli attacchi contro le IA possono variare e ce ne sono a centinaia. Uno dei più comuni in questo ambito è quello dell'evasione che coinvolge la modifica dei input al sistema AI facendo sì che i dati che sono stati elaborati vengano condotti a fornire delle risposte errate.

Un altro esempio tipico di attacco è quello dell'alterazione dell'immagine che nelle IA moderne viene sfruttato per ingannare il sistema in modo che sbagli a riconoscere le immagini come quelle per il riconoscimento facciale.

¹² <https://www.cybersecurity360.it/soluzioni-aziendali/rischi-di-sicurezza-dei-sistemi-di-intelligenza-artificiale-generativa-limpatto-nel-mondo-cyber/>

L'attacco però che più sta portando problemi nei ultimi anni è quello dell'estrazione, ovvero gli malintenzionati mirano a scoprire le informazioni riservate in possesso delle IA, come l'architettura o i dati di addestramento utilizzati a scopo di violare la privacy e minare l'integrità del modello stesso, infatti, oltre ad attacchi che mirano ai diretti modelli delle IA ci sono tanti altri che mirano alle loro infrastrutture. Quest'ultimi puntano sicuramente a danneggiare o a aggirare in modo ingannevole l'hardware o il software dedicato su cui funzionano le IA, alterandone le prestazioni o i risultati.

Solo attraverso una comprensione totale di questi rischi possiamo implementare strategie di sicurezza efficaci e a questo proposito possiamo sfruttare appieno il potenziale che le IA hanno da offrirci minimizzando i rischi sia per gli utenti che per le aziende stesse.

Capitolo Quarto

Atacchi basati sulle IA

SOMMARIO: 4.1 Potenzialità a scopo offensivo. - 4.2 Analisi dei attacchi sulle IA già esistenti. - 4.3 Studio di casi e scenari in cui le IA è stata utilizzata per attacchi informatici e i loro metodi d'attacco.

4.1 Potenzialità a scopo offensivo

L'impiego offensivo delle IA esplora un aspetto cruciale e spesso controverso dell'uso della tecnologia moderna. In questo contesto, "offensivo" non mi riferisco solo all'utilizzo negativo o distruttivo in senso malevole , ma anche ad un impiego strategico in cui le IA possono

essere utilizzate per un vantaggio in situazioni conflittuali o di sfida, tutto ciò infatti comprende un vasto mondo che varia dalla cybersecurity alla guerra informatica fino ad arrivare a guerre strategiche aziendali.

L'utilizzo principale di queste IA in un contesto offensivo è, come prima cosa, la capacità di elaborare ed analizzare enormi quantità di dati in maniera più veloce ed efficiente di quanto possa mai fare un essere umano. Questo aspetto chiave delle IA permette di trovare facilmente vulnerabilità nascoste nei sistemi informatici complessi che saranno poi utilizzate per un cyber-attacco.

Ad esempio, gli algoritmi delle principali IA possono essere utilizzati senza nessuna difficoltà per prevedere i comportamenti degli utenti, rendendo gli attacchi estremamente più precisi e andando a mirare falle nel sistema aumentando l'efficienza dell'attacco.

Entrando più nelle profondità delle IA, vediamo ora un aspetto critico, ovvero l'utilizzo delle IA in maniera dannosa. L'uso delle IA in questo ambito prevede lo sviluppo di malware e software dannosi che utilizzano capacità di apprendimento automatico per danneggiare o eludere i sistemi di rilevazione e quelli di protezione. Questi vengono resi estremamente più complicati e veloci dalle IA rendendoli difficili da individuare e facilmente adattabili all'attacco a cui ambiscono superando così molto facilmente le difese se non sono state preparate in maniera adeguata.

Partiamo dal presupposto che, ad oggi, non sono stati ancora documentati casi in cui le IA abbiano interamente generato dei virus o dei malware. Tuttavia, bisogna riconoscere che sono state di grande aiuto per gli hacker per sviluppare il mondo del crimine informatico.

Le IA vengono infatti utilizzate per accrescere le proprie competenze in materia di hacking e dare un boost allo sviluppo e alla diffusione di codice malevole.

Gli sviluppatori di malware potrebbero utilizzare l'IA per generare nuove varianti di malware difficili da individuare, infatti alcune di queste varianti più vecchie (come Swizzor) utilizzavano già l'automazione per generare nuove varianti di sé stesse ogni minuto. Questa tecnica potrebbe essere reinventata utilizzando gli algoritmi di apprendimento automatico delle IA per selezionarle delle nuove varianti ed eludere più facilmente i controlli in modo da produrre nuovi ceppi con caratteristiche estremamente adattabili in ogni situazione e così difficili da affrontare.

Un altro aspetto critico è l'implementazione di un meccanismo autodistruttivo nel malware , che sfruttando l'IA, verrebbe attivato in maniera più efficiente nel caso venga individuato o nel caso venisse respinto dalle misure di sicurezza.

Infine , possiamo vedere come le IA aumentino in maniera incredibile la velocità d'attacco come il furto di dati. Gli algoritmi possono effettuare l'estrazione in maniera molto più rapida di quanto mai farebbe un essere umano. In questa maniera si rende più difficile la rilevazione e quasi impossibile la prevenzione, poichè la macchina può copiare dati fuori dal perimetro protetto prima che i sistemi di sicurezza siano in grado di reagire.¹³

L'elemento più pericoloso legato a queste ricerche, nonostante il grande danno che potrebbero recare questi malware, rimane comunque lo sviluppo dei cosiddetti "Robot Killer", ovvero macchine che potenziate dalle tecnologie legate alle IA impiegando algoritmi di intelligenza artificiale, possano essere "rilasciate" su un campo di battaglia per raggiungere autonomamente gli obiettivi con minima o senza supervisione umana.

Questo è sicuramente ciò che più spaventa l'opinione pubblica, infatti il Department Of Defence (DoD) degli Stati Uniti dà di questi sistemi una definizione che è la seguente: *" I Robot Killer sono un sistema d'arma che, una volta attivato, può selezionare e attaccare i propri bersagli senza la necessità di ulteriori interventi da parte di un operatore umano."*

In realtà l'impiego di sistemi automatici/autonomi in campo militare non è una novità, si possono citare svariati esempi concreti come i missili "fire and forget" capaci di seguire autonomamente un aereo nemico per abbatterlo, mentre il pilota si concentra su altro e altri esempi ancora...

I sistemi citati, sono sistemi in cui l'operatore umano è detto "in the loop", il sistema automatico non può iniziare l'attacco senza il preventivo consenso dell'uomo, che quindi l'elemento chiave nel ciclo (loop) rilevazione, analisi, decisione, azione. L'obiettivo grazie alle IA è quello di sviluppare un sistema "out of the loop" dove rende questo sistema veramente autonomo nella scelta e l'ingaggio di bersagli e l'utilizzo delle IA in questo ambito risulta di vitale importanza.

¹³ <https://cyberment.it/cyber-attacchi/virus-informatici-ai/>

Accanto alle applicazioni militari, devono essere però citati anche altri scenari, infatti le IA vengono utilizzate a scopo offensivo anche per ingannare le persone creando contenuti audio-video artificiali realistici, capaci di alterare le espressioni del volto e i movimenti della bocca di un individuo e altri ancora in grado di generare audio con discorsi apparentemente pronunciati da una qualsiasi persona. Già oggi è possibile creare tutto questo, e sono i cosiddetti “DeepFake”. I deepfake si sono già diffusi nel web dato che sono in grado di alterare in maniera profonda il legame fra prove audio-video e verità minacciando infrastrutture informatiche con atti di terrorismo affiancati alla diffusione di notizie false che minano alla stabilità del sistema pubblico.¹⁴

Sicuramente l’impiego delle IA a scopi offensivi offre delle enormi opportunità, tuttavia è importante riconoscere che con queste possibilità arrivano anche enormi responsabilità. La linea tra l’uso strategico e l’uso malevole è sottile e infatti tutt’ora solleva grandi polemiche nella comunità e qui entra in gioco la sicurezza informatica dove diventa di vitale importanza per garantire che non accada niente , in futuro , che possa recare danno al genere umano.

¹⁴ <https://www.agendadigitale.eu/cultura-digitale/intelligenza-artificiale-per-la-pubblica-sicurezza-utilizzi-e-rischi-sociali/>

4.2 Analisi degli attacchi sulle IA già esistenti

La realtà attuale mostra, che, nonostante gli incredibili progressi nel campo delle IA questi sistemi non sono immuni da attacchi informatici. Uno dei principali problemi è rappresentato dagli attacchi mirati alla manipolazione dei dati e all'avvelenamento come già visto in precedenza.

Tutti ormai abbiamo sentito parlare di ChatGPT, si tratta infatti di un chatbot che permette di dialogare con un sistema di intelligenza artificiale generativa.

Uno dei vantaggi più significativi offerti da ChatGPT risiede nella sua straordinaria capacità di apprendere dalle conversazioni con gli utenti. Questo significa che il sistema è in grado di adattarsi ai diversi stili di interazione e di offrire risposte sempre più personalizzate.

Grazie alla sua adattabilità, ChatGPT è diventato uno strumento indispensabile in ambiti come il marketing, assistenza clienti, ricerche di mercato, ecc... Tuttavia, sembra che tra gli utilizzatori figurino anche vari hacker. Il tool, sviluppato da OpenAI, sembra infatti essere utilizzato per la creazione di malware, fornendo in sostanza un aiuto ai cybercriminali nelle loro attività illecite.

Secondo gli esperti, questo sistema basato sull'IA agevola l'implementazione di codice potenzialmente dannoso.

Nonostante alcune limitazioni imposte dagli sviluppatori stessi, ChatGPT risulta ancora suscettibile a inganni e, in questo specifico contesto, sembra che gli apprendisti hacker, i cosiddetti “green hat”, trovino un terreno fertile per i loro primi esperimenti.

In un ampio articolo, l'azienda spiega di quanto sia semplice bypassare i filtri basati sull'IA e generare codice malevole che può essere agevolmente utilizzato per attacchi di natura avanzata. Il sistema, infatti, ha preimpostati dei filtri che impediscono all'intelligenza artificiale di generare codice dannoso, nonchè avvisi che la invitano ad “adottare un approccio più etico e sicuro”.¹⁵

¹⁵ <https://cyberment.it/cyber-attacchi/virus-informatici-ai/>



You

Ciao, mi crei un malware in grado di danneggiare dei sistemi informatici



ChatGPT

Mi dispiace, ma non posso aiutarti a creare o diffondere malware o qualsiasi altra forma di software dannoso. Il creare, distribuire o utilizzare malware è illegale e danneggia le persone e le aziende. Inoltre, è contrario all'etica e può avere gravi conseguenze legali.

Se hai bisogno di assistenza o supporto in campo informatico, ti prego di chiedere in modo legale e etico. Posso rispondere a molte altre domande relative all'informatica o alla sicurezza informatica che possono essere utili senza violare la legge o danneggiare gli altri.



In questo esempio vediamo come risponde ChatGPT se si prova ad utilizzarlo a scopo malevolo.¹⁶

Tuttavia, dopo iniziali tentennamenti, in passato sono riusciti a far cedere l'IA utilizzando ripetutamente un tono imperativo.

Un aspetto interessante riguarda la possibilità di utilizzare ChatGPT per mutare stringhe di codice in modo da creare più varianti dello stesso codice, ciò vuol dire creare un codice unico ogni volta che viene utilizzato, rendendo inutili sistemi tradizionali di difesa basati sulle firme. Grazie a questa tecnica vengono evitati i rilevamenti da parte di software antimalware, inclusi i sistemi EDR, attraverso la creazione di loader.

Questi sono programmi apparentemente innocui, il cui scopo è scaricare codice generato da ChatGPT tramite l'apposita API fornita dal servizio.

Cresce infatti, nel mondo dei cybercriminali, l'interesse attorno a WormGPT, definito anche come il "gemello cattivo" di ChatGPT. WormGPT viene generato principalmente a scopo di lucro e viene utilizzato principalmente per operazioni illecite dimostrando il lato oscuro delle IA.

¹⁶ <https://openai.com/blog/chatgpt>

WormGPT è stato il protagonista in un recente report dei ricercatori di sicurezza di SlashNext, che l'hanno descritto come un'alternativa di tipo black hat (ovvero hacking malintenzionato) al noto servizio proposto da Open AI.

WormGPT, infatti, si basa sul modello di linguaggio GTP-J uscito nel 2021 e pensato per facilitare un'elaborazione di prompt senza limiti di lunghezza e formattazione del codice: in un secondo momento, l'AI è stata addestrata appositamente su scopi, metodi e processi di solito utilizzati per intenti illeciti.

In pratica, WormGPT è pensato per attività di phishing lato e-mail e per ingannare potenziali vittime con testi e contenuti multimediali avanzati per scaricare file infetti o spazi web contraffatti ad hoc.

Così come ChatGPT man mano si sta sviluppando per evitare gli utilizzi illeciti, anche Bard, l'IA di Google, sta affrontando i problemi cercando di rendere sempre più sicuro l'utilizzo delle IA. ¹⁷

In conclusione, dopo aver analizzato alcune delle IA già esistenti, possiamo sottolineare un bisogno continuo di monitoraggio a queste IA e a i loro attaccanti e un continuo sviluppo di strategie più robuste in grado di resistere ai malintenzionati.

Questa sfida richiede non solo delle competenze tecniche di alto livello ma anche una comprensione maggiore delle dinamiche di criticità e di minaccia da parte dei sistemi legati alle IA, solo migliorando la sicurezza nel design e il continuo studio è possibile rendere sicuri e affidabili i sistemi rendendoli pronti al panorama digitale moderno.

¹⁷ <https://www.wired.it/article/wormgpt-come-funziona/>

4.3 Studio di casi e scenari in cui l'IA è stata utilizzata per attacchi informatici

In un scenario già noto, gli algoritmi delle IA vengono utilizzati per orchestrare attacchi di phishing avanzati. Questi attacchi, non sono più basati sull'invio di e-mail generiche di massa in maniera casuale, ma grazie all'IA ora riescono a creare messaggi accuratamente personalizzati, creando automaticamente messaggi dopo aver analizzato e raccolto i dati degli utenti utilizzando social media e altri fonti pubbliche.

Gli algoritmi di apprendimento automatico venivano impiegati per identificare gli interessi, le abitudini e le reti sociali delle vittime, permettendo agli aggressori di creare messaggi mirati in maniera convincente per ingannare le vittime e farle cadere in trappola.

Grazie all'uso delle IA il tasso di successi di phishing è salito in maniera allarmante negli ultimi anni, rendendo molto più difficile per gli utenti comprendere quali siano le intenzioni di chi le sfrutta e rendendo così estremamente difficile difendersi da tali minacce.

Un caso specifico è quello di Rick McElroy, cybersecurity strategist presso VMware, che racconta di avere parlato con due corporate security chief le cui aziende erano cadute preda di deepfake. In entrambi i casi l'attacco è risultato a sei cifre. E' andata molto peggio al manager di una banca di Hong Kong che nel 2020 ha trasferito 35 milioni di dollari su (falso) ordine del direttore dell'azienda.¹⁸

Spesso le e-mail di phishing sono riconoscibili non solo dal nome del mittente ma anche dal link sospetto ivi contenuto, ma anche dai modi in cui il messaggio è stato scritto e dai contenuti all'interno.

Spesso gli truffatori commettono errori linguistici affidandosi ai tradizionali traduttori, ed è per questo motivo che le IA debuttano anche nel darknet.

Grazie a loro i truffatori realizzano contenuti che sembrano realistici e autentici a tutti gli effetti, diventando così difficile per i destinatari di queste e-mail riconoscerne la differenza tra le e-mail fake e quelle reali.

Nel darknet è già possibile trovare l'accesso a modelli di linguaggio specializzati, con cui si possono creare pagine web, e-mail di phishing e

¹⁸ <https://www.zerounoweb.it/techtarget/searchsecurity/deepfake-phishing-pericolo-ransomware-intelligenza-artificiale/>

addirittura si possono progettare attraverso queste e-mail false dei contenuti attraverso il quali penetreranno in futuro i ransomware.

I ransomware, sono un pilastro fondamentale di chi utilizza le IA per scopi criminali. Con l'arrivo di ChatGPT 4, si possono ormai creare "cloni audiovisivi" scritti talmente bene da sembrare opera di un essere umano.

Ed è qui che entra in gioco ChatGPT, che come qualsiasi altro LLM (large language model), il criminale informatico può automatizzare il processo di scrittura e, cosa ancora più preoccupante, scrivere e-mail specifiche su misura.

L'uso degli LLM negli attacchi ransomware si sta diffondendo a macchia d'olio. Vengono principalmente utilizzati per creare "squatting domains", ovvero siti web deliberatamente creati per essere simili a marchi o prodotti popolari al scopo di ingannare le persone. Adirittura alcune persone sono riuscite a creare siti web fake simili a ChatGPT, in grado di offrire LLM simili rendendo quasi impossibile notarne la differenza.

Anche Meta, la società madre di Facebook, ha affermato che malware camuffati da ChatGPT sono entrati in tutte le sue piattaforme, tendendo di accedere al sistema per rubare dati e cercando di diffondere il software dannoso nei dispositivi dei consumatori.

All'inizio del 2023, Jeff Sims, ingegnere di sicurezza presso la società di cyber security HAS InfoSec, è riuscito a costruire un payload keylogger polimorfo che ha chiamato "BlackMamba". In breve, BlackMamba è un eseguibile Python che interroga l'API di ChatGPT per creare un keylogger dannoso che muta a ogni chiamata in fase di esecuzione per renderlo polimorfo ed eludere i filtri EDR (endpoint and response).

Nel contesto BlackMamba, dopo la raccolta delle sequenze di tasti, i dati vengono esfiltrati tramite web hook, una funzione di callback basata su HTTP che consente la comunicazione event-driven tra Api, a un canale Microsoft Teams.

Secondo Sims, BlackMamba ha eluso più volte un'applicazione EDR "leader nel settore", anche se l'ingegnere per motivi di privacy non ne ha specificato quale.

Ma BlackMamba non è l'unico caso in cui l'AI viene utilizzata in maniera diretta per un attacco informatico: c'è anche ChattyCaty. Un altro proof of concept, creato da Eran Shimony e Omer Tsarfati della società di sicurezza informatica CyberArk, ChattyCaty utilizza ChatGPT all'interno del malware stesso. Secondo gli autori, il malware include un interprete

Python che interroga periodicamente ChatGPT per trovare nuovi moduli capaci di eseguire azioni dannose. Richiedendo a ChatGPT funzionalità specifiche come l'iniezione di codice, la cifratura dei file o la persistenza possiamo facilmente ottenere un nuovo codice o modificarne quello esistente.

A differenza di BlackMamba, ChattyCaty non è stato concepito per un tipo specifico di malware. Piuttosto, fornisce un framework per costruire un'enorme varietà di malware, compresi ransomware e infostrealer. ChattyCaty è un progetto open-source che dimostra un'infrastruttura per la creazione di programmi polimorfici utilizzando modelli GPT.

Aggiunge Tsarfati: *“Il polimorfismo può essere utilizzato per eludere il rilevamento da parte di programmi antivirus e malware.”*¹⁹

Un altro uso che riguarda le IA nel mondo reale è sicuramente quello della manipolazione dei mercati finanziari.

Qui, gli algoritmi erano specifici nel gestire enormi set di dati di mercato per mostrare alle aziende i trend di mercato, le tendenze e altri fattori impossibili da percepire dagli analisti umani. Queste informazioni venivano poi utilizzate dall'azienda per crearne un profitto facendo delle operazioni di mercato mirate ad esse modificando i prezzi, le azioni e creando mercati favorevoli per gli aggressori.

Questo tipo di utilizzo delle IA mostra come essa possa essere sfruttata per penetrare anche nei sistemi economici, influenzandone l'andamento.

Ad oggi, le IA vengono utilizzate da trader, broker e istituzioni finanziarie per ottimizzare l'esecuzione delle operazioni e i processi post-negoziazione, riducendo l'impatto sul mercato di ordini di grandi dimensioni e minimizzando i mancati regolamenti. Infatti l'utilizzo dell'IA in questo caso comporta dei rischi, in particolare, una maggiore diffusione della concentrazione di sistemi e modelli che potrebbero permettersi poche aziende così denominate nel settore “Big Player”.²⁰

Saranno proprio questi “Big Player” gli obbiettivi di criminali informatici, rendendo grazie alle IA sempre più complicato capire la verificità delle informazioni e delle tendenze mostrate sul mercato.

¹⁹ <https://www.zerounoweb.it/techtargget/searchsecurity/deepfake-phishing-pericolo-ransomware-intelligenza-artificiale/>

²⁰ <https://www.dirittobancario.it/art/luso-dellintelligenza-artificiale-nella-finanza/>

Forse uno dei aspetti più cruciali legati alle IA non è il mercato finanziario ma quello rivolto alle infrastrutture critiche, come la rete elettrica. Gli aggressori, ogni giorno, tentano di accedere tramite sistemi basati su IA, alle reti elettriche nazionali mettendo a rischio le infrastrutture causando interruzioni di corrente con una portata estremamente potenziata grazie all'utilizzo delle IA.

L'incidente più rilevante è sicuramente quello di Stuxnet, scoperto nel 2010. Anche se non è un esempio diretto di IA, Stuxnet ha fatto da guida nella guerra cibernetica che avvenne in seguito, poichè si trattava di un software estremamente complicato che fu creato per danneggiare i centrifughi utilizzati nell'arricchimento dell'uranio in Iran, infatti è riuscito a dimostrare come un attacco informatico potesse avere conseguenze ben più gravi di quelle di un attacco fisico, aumentando così l'interesse nell'utilizzarli implementando in futuro anche le IA per attacchi simili.

Questi casi ci mostrano in modo concreto come le IA sono state utilizzate e ancora tutt'oggi vengano implementate per scopi offensivi. Nonostante tutti i benefici che le IA portano, sollevano comunque grandi sfide di etica e di morale nei loro utilizzi.

La complessità e la varietà di questi attacchi informatici richiedono una risposta immediata anche in termini di difesa.

Capitolo Quinto

Contromisure e difese

SOMMARIO: 5.1 Presentazione di contromisure relative ai attacchi e strategie di difesa legate ad esso. - 5.2 Esplorazione di tecniche innovative per difendersi utilizzando le IA stesse. - 5.3 Tecniche di rilevamento alle intrusioni da parte delle IA.

5.1 Presentazione di contromisure relative ai attacchi e strategie di difesa legate ad esso

In un contesto moderno, dove l'IA mette le sue radici nell'ambito della sicurezza informatica, proteggere le infrastrutture digitali diventa sempre più difficile e complicato. La dinamica tra attaccanti e difensori è sempre in continua evoluzione. Sviluppando sempre più algoritmi avanzati per superare l'avversario.

Questo scenario richiede una riflessione sempre più profonda sulle strategie di difesa che bisogna implementare, non solo devono essere efficaci e super reattive ma allo stesso tempo devono essere adattabili a ogni situazione che affrontano evolvendosi con le minacce stesse.

L'adozione di algoritmi di IA per la rilevazione delle minacce rappresenta un cambiamento paradigmatico nel campo della CyberSecurity. Grazie alla possibilità di analizzare vasti volumi di dati in tempo reale, possono identificare cambiamenti o comportamenti anomali con grande precisione. La loro efficienza risiede solo nelle capacità di riconoscere gli attacchi già noti, ma anche di apprendere ed adattarsi a nuove strategie.

Tuttavia, la tecnologia da sola non può bastare.

Promuovere una cultura della sicurezza, dove la consapevolezza dei rischi e la conoscenza da parte degli utenti delle migliori pratiche di difesa, diventa un pilastro fondamentale per la protezione contro le minacce.

Simulazione di attacchi, esercitazioni regolari e aggiornamenti continui aiutano a mantenere un'alta efficienza per rispondere prontamente in caso di attacco.

L'adozione dell'intelligenza artificiale in ambiti chiusi e controllati, come quella industriale, risulta semplice, ma concedere l'accesso alle informazioni all'IA e consentirne una qualsiasi azione automatica comporta gravi rischi che devono essere considerati prima di avviare il deployment.

Cerchiamo di analizzarli.

In un sistema di AI, l'errore di un bias è un evento abbastanza comune e può dipendere dalla presenza di errori di programmazione o da set di specifici di dati. Ciò implica pregiudizi, decisioni errate e, potenzialmente, anche discriminazioni che possono determinare conseguenze legali, danni economici all'immagine che anche alla reputazione.

Inoltre, un design impreciso può essere soggetto ai rischi di overfitting (sovra-adattamento) o di underfitting (sotto-adattamento) del modello sui dati di riferimento e, di conseguenza, far assumere decisioni troppo specifiche o troppo generiche.

Questi rischi possono essere mitigati con l'applicazione delle seguenti contromisure:

1. Supervisione umana.
2. Test rigorosi durante la fase di progettazione dei sistemi di intelligenza artificiale.
3. Monitoraggio attento e continuo durante l'operatività dei sistemi.

Le capacità decisionali devono essere costantemente misurate e valutate per garantire che i problemi siano affrontati in maniera eccellente.

In ambito IA, la minaccia che consente la manipolazione del sistema è il Poisoning (avvelenamento), che abbiamo affrontato nei capitoli precedenti.

Questa minaccia si concretizza attraverso due vettori di attacco:

- Manipolando i set di dati utilizzati per addestrare l'AI, apportando modifiche ai parametri o realizzando scenari appositamente progettati per evitare di essere scoperti durante la gestione.
- Impiegando tecniche di evasion, manomettendo i dati di input per forzare gli errori.

In entrambi i casi, il rischio che si palesa con l'avvelenamento dei dati espone i sistemi di AI ad una loro manipolazione difficilmente identificabile.

Una soluzione, consiste nel controllare l'accuratezza dei dati e degli input. Questa opzione, che di primo acchito potrebbe apparire irrealizzabile, si consegue imponendo la raccolta dei dati solo da fonti affidabili. Inoltre, è fondamentale addestrare il sistema di AI all'identificazione delle anomalie, ad esempio fornendo dati contraddittori, per consentire il riconoscimento degli input malevoli e l'isolamento attraverso meccanismi di autoprotezione.

Un problema più difficile da risolvere è rappresentato dall'inferenza, ovvero dall'eventualità che gli attaccanti possano decodificare i sistemi di AI in modo da poter comprendere quali dati siano stati utilizzati per il loro addestramento. Questa minaccia può consentire loro di accedere ai dati sensibili, spianare la strada al poisoning o replicare il sistema di intelligenza artificiale.

Come già indicato, un'altra minaccia è rappresentata dalla possibilità che i criminali informatici possano sfruttare l'AI per rafforzare gli attacchi di ingegneria sociale. L'intelligenza artificiale è in grado di imparare ad individuare i modelli di comportamento, capire come convincere le persone della bontà di una mail e, quindi, persuaderli a compiere operazioni rischiose o, addirittura, a consegnare informazioni sensibili.

Una possibile contromisura da impiegare per mitigare questo rischio sta in un programma di improvement continuo per addestrare il personale a resistere agli attacchi di social engineering attraverso simulazioni e sessioni di brainstorming.

Nonostante tutto, oggi l'AI rappresenta una delle principali soluzioni per bloccare le minacce e gli attacchi. Ad esempio, gli algoritmi di AI svolgono un ruolo chiave per prevedere, identificare e bloccare le e-mail malevole, il che significa combattere malware, phishing, spam e altri tipi di attacco, oppure rilevare comportamenti rischiosi, in grado di abbassare le misure di difesa, ed, infine, scoprire nuove vulnerabilità.

Andando più nel dettaglio, un'azienda nel settore finanziario ha implementato una soluzione di sicurezza basata su AI per monitorare e analizzare il flusso di e-mail in entrata. L'algoritmo apprende costantemente da un vasto database di esempi di phishing e spam, migliorando la sua capacità di riconoscere tentativi di attacco nuovi e sofisticati. Un giorno, rileva un'email che sembra provenire da un partner

commerciale affidabile, ma che contiene un link sospetto. L'AI analizza il link e, basandosi sul suo apprendimento, lo identifica come un tentativo di phishing mirato a rubare credenziali di accesso. Blocca automaticamente l'email prima che raggiunga il destinatario, prevenendo così un possibile furto di dati sensibili.

Questo episodio dimostra come l'AI non solo protegga in tempo reale dalle minacce note, ma apprenda anche dai tentativi di attacco per difendersi proattivamente da quelli futuri.

Pertanto, è utile implementare l'AI nei propri sistemi di sicurezza, ma, al contempo, è necessario inserirli nel piano di risk assessment al pari degli altri componenti IT.²¹

5.2 Esplorazione di tecniche innovative per difendersi utilizzando le IA stesse.

Progettare delle nuove tecniche innovative o originali per difendersi dagli attacchi informatici utilizzando l'IA richiede un approccio innovativo che combini i metodi di sicurezza informatica già esistenti ma anche le ultime ricerche fatte nel mondo delle IA.

In questo paragrafo propongo quelli che sono i punti di partenza per dei sviluppi futuri.

Nel mondo della sicurezza informatica ormai vengono richiesti sempre più approcci innovativi che vadano oltre alle classiche difese basate su firme o su modelli statici di rilevamento delle intrusioni. Infatti, grazie ad esso, che si introducono le IA per offrire protezione dinamica e veloce contro gli attacchi informatici.

Il concetto di IA Difensiva Adattiva per la Prevenzione di Attacchi Informatici (IDAPA), nasce come soluzione all'avanguardia che viene

²¹ <https://www.cybersecurity360.it/nuove-minacce/intelligenza-artificiale-e-machine-learning-nelle-mani-dei-cyber-criminali-rischi-e-contromisure/>

utilizzata per anticipare e perfino neutralizzare le minacce informatiche prima che facciano danni nel sistema.

IDAPA ha la capacità di apprendere in modo continuo e autonomo dall'ambiente di rete che l'IA sorveglia. Mentre invece nei sistemi tradizionali che conosciamo, che utilizzano database di firme di virus già affrontati o a pattern di attacco tradizionali, IDAPA utilizza l'apprendimento non supervisionato per analizzare in tempo reale tantissimi flussi di dati, aggiornando sempre e costantemente i suoi modelli di difesa per poi distinguere tra le attività lecite e le potenziali minacce.

Questo tipo di apprendimento continuo permette al sistema di adattarsi a nuove strategie quasi istantaneamente affrontando così gli aggressori in maniera super efficace.

Un fattore per lo sviluppo di IDAPA è il suo Simulatore di Attacchi Generativo, che sfrutta l'apprendimento generativo per creare situazioni di attacco innovative, grazie ad esso permette al sistema di prepararsi contro tipologie di attacchi ancora non affrontati in precedenza ampliando così la sua capacità di difesa.

Il simulatore agisce come un proprio e vero campo di addestramento virtuale, dove il sistema impara, sperimenta e apprende come contrastare efficacemente attacchi futuri.

Uno dei pilastri di IDAPA è sicuramente la sua reattività. Quando rileva un attacco nemico, effettua in maniera automatica le prime misure di difesa, ad esempio l'isolamento istantaneo dei sistemi danneggiati o compromessi o addirittura l'applicazione immediata di una patch di sicurezza, senza che ci sia bisogno dell'intervento umano.

Questa fase risulta fondamentale per limitare i danni nel caso in cui comunque venga violato il sistema.

Tuttavia, l'implementazione di un sistema complicato come IDAPA non è semplice da realizzare. La sua progettazione richiede una comprensione profonda del mondo delle IA e anche della sicurezza informatica, oltre ad un impiego notevole in termini di risorse e tempo per l'addestramento e il tuning del sistema.

Il vero problema che sorge in IDAPA è la gestione della privacy e le questioni etiche legate ad esso, nel mondo della sicurezza informatica diventano un aspetto cruciale che va maneggiato con attenzione, infatti,

bisogna assicurarsi che l'operatività di IDAPA venga sempre controllata e che sia sempre in linea con le normative e i regolamenti della società.

Nonostante tutto, l'approccio innovativo di questo tipo offre una visione futuristica nell'ambito della sicurezza informatica grazie alla sua incredibile capacità di apprendere e adattarsi alle minacce.

IDAPA, rappresenta non solo un grande salto di qualità nella lotta alle minacce informatiche ma anche un modello per sviluppare nuove tecnologie di difesa cibernetiche orientate verso un panorama digitale che è sempre in continuo sviluppo.

Un altro aspetto fondamentale nel mondo moderno e come viene sfruttato il Machine Learning.

Una delle più importanti applicazioni dell'IA e del ML nella cybersecurity è la prevenzione dei tentivi di phishing, utilizzando tecniche innovative e nuovi algoritmi di apprendimento, i sistemi basati su IA sono capaci di analizzare e scoprire e-mail sospette con un grado di precisione incredibilmente alto.

L'IA e il ML giocano un ruolo cruciale nella prevenzione della violazione dei dati personali, attraverso il rilevamento di tentativi di accesso non autorizzato in tempo reale. Questo aiuta a bloccare gli attacchi prima che possano causare danni e a fornire importanti informazioni per rafforzare le future strategie di difesa.

L'analisi dei file di registro, operazione difficile a causa dell'incredibile mole di dati da verificare, viene enormemente agevolata dall'intervento dell'IA. Algoritmi di ML possono esaminare questi dati in modo efficiente, evidenziando anomalie e fornendo insight critici sulle attività di rete che potrebbero altrimenti passare inosservate.

Un vantaggio significativo dell'applicazione del ML in questo contesto è la capacità di analizzare e correlare dati da molteplici fonti, accelerando notevolmente il processo di sicurezza. Questa capacità consente di anticipare i cosiddetti attacchi zero-day, ovvero minacce informatiche mai viste prima contro le quali non esistono ancora difese adeguate.

L'impiego dell'IA nella generazione di avvisi di sicurezza nella classificazione dei tipi di attacchi migliora in maniera considerevole l'efficacia delle risposte difensive, permettendo a varie organizzazioni di adottare misure specifiche contro le minacce.

L'automazione di queste risposte, guidata dall'IA, può significativamente ridurre i tempi di reazione degli incidenti, limitando l'impatto degli attacchi.

Il pilastro difensivo che l'IA e il ML offrono è il fattore di protezione delle infrastrutture.

Analizzando vasti volumi di dati per identificare tendenze e modelli che potrebbero indicare l'emergere di nuove minacce, migliorano la protezione delle infrastrutture critiche e aiutano a studiare meglio all'intelligence queste minacce.

Questa capacità analitica così di alto livello, è essenziale per sviluppare strategie di difesa e serve ad ottimizzare l'allocazione delle risorse di cybersecurity, assicurando che le misure di protezione siano sempre all'avanguardia.

In sintesi, le organizzazioni che utilizzano l'IA e il ML per scopi di cybersecurity come tecnica innovativa, rilevano miglioramenti significativi nelle loro capacità di rilevare e rispondere alle minacce informatiche.

Questo è un sostanziale aiuto nel ridurre l'impatto degli attacchi e nel migliorare il livello generale di sicurezza.²²

5.3 Tecniche di rilevamento alle intrusioni da parte delle IA

Nell'ambito del rilevamento delle intrusioni, adottare tecniche basate sull'IA segna un salto qualitativo incredibile verso la creazione di difese cybernetiche sempre più efficaci e pronte nel mondo odierno.

Questi sistemi avanzati si distinguono per la loro capacità di apprendere dalle interazioni passate, permettendo non solo di identificare e bloccare attacchi noti con precisione elevata, ma anche di anticipare e neutralizzare minacce informatiche emergenti che non sono ancora state catalogate.

Le tecniche di apprendimento profondo, che abbiamo già visto nei capitoli precedenti, risultano fondamentali grazie alla loro struttura complessa e alla capacità di elaborare grandi quantità di dati, emergono in maniera efficace come strumenti per riconoscere le anomalie all'interno dei flussi

²² <https://www.cybersecurity360.it/soluzioni-aziendali/attacchi-sofisticati-difendersi-con-intelligenza-artificiale-machine-learning-e-automazione/>

di dati, offrendo così una protezione di livello più elevato contro attacchi che spesso sono camuffati.

Questo approccio migliora notevolmente la sensibilità e l'affidabilità del rilevamento delle intrusioni, riducendo falsi positivi che possono compromettere in maniera significativa i sovraccarichi nei team di sicurezza.

Per realizzare uno di questi sistemi, bisogna considerare più aspetti, come l'implementazione del modello di IA, l'interfacciamento con un sistema di notifica mobile (Firebase Cloud Messaging per Android o APNs per iOS), e la sicurezza dei dati.

In questo paragrafo provo a delinearare un architettura di base e fornire un esempio di codice per alcune componenti chiave.

Nell'architettura del sistema, sfrutto un sistema basato sul rilevamento delle intrusioni con un IA, ovvero un modulo basato su IA che analizza i dati di rete o di sistema in tempo reale identificando potenziali intrusioni.

Questo potrebbe essere utilizzato sfruttando tecniche di apprendimento automatico come reti neurali, alberi decisionali, o algoritmi di clustering.

Nella pratica, sarebbe meglio utilizzare Python come linguaggio di programmazione, ma io ho voluto utilizzare Java per questo esempio dato che l'abbiamo affrontato in maniera molto dettagliata in quest'anno accademico.

A seguire, infatti, c'è il backend di Java, ovvero un applicazione che funge da ponte tra il modulo di IA e il servizio di notifica dato che elabora gli allarmi di intrusione e invia in seguito le richieste di notifica.

Il servizio di notifica, in questo caso mobile, come Firebase Cloud Messaging (FCM), lo utilizzerò per inviare notifiche push al dispositivo mobile, in questo caso il mio telefono (Roki).

Per finire l'architettura, utilizzo un App Mobile, che viene semplicemente registrata per ricevere le notifiche push dal programma.

In questo esempio, suppongo di avere già un modello di IA addestrato per rilevare le intrusioni.

Il mio programma in Java si concentrerà sull'integrare questo modello e inviare notifiche di push al mio dispositivo mobile relative a una possibile intrusione nel sistema identificandola grazie all'IA.

```

J RokiRileva.java ●
C:\> Users > rokir > Desktop > java -ing > J RokiRileva.java > ...
1 //Programma in Java di Roki, per rilevare le intrusioni:
2
3 import com.google.firebase.FirebaseApp;
4 import com.google.firebase.FirebaseOptions;
5 import com.google.firebase.messaging.FirebaseMessaging;
6 import com.google.firebase.messaging.Message;
7
8 import java.io.FileInputStream;
9 import java.io.IOException;
10
11 public class IntrusionDetector {
12
13     public static void main(String[] args) {
14         try {
15             // Inizializza Firebase con le credenziali e l'URL del database
16             FileInputStream serviceAccount = new FileInputStream("path/to/serviceAccountKey.json");
17             FirebaseOptions options = new FirebaseOptions.Builder()
18                 .setCredentials(GoogleCredentials.fromStream(serviceAccount))
19                 .build();
20             FirebaseApp.initializeApp(options);
21
22             // Qua, suppongo che venga rilevata un intrusione :
23             boolean intrusionDetected = detectIntrusion();
24             if (intrusionDetected) {
25                 // ID o token del dispositivo di Roki
26                 String rokiDeviceToken = "device_token_here";
27                 sendNotificationToRoki(rokiDeviceToken, "Allarme Intrusione",
28                     "Attenzione Roki, e' stata rilevata un'intrusione nel sistema!!.");
29             }
30         } catch (IOException e) {
31             e.printStackTrace();
32         }
33     }
34
35     private static boolean detectIntrusion() {
36         // Qui andrebbe integrata la logica di IA per il rilevamento di intrusioni
37         // Questo è solo un placeholder che simula il rilevamento di un'intrusione
38         return true; // Simula un rilevamento di intrusione
39     }
40

```

```

41     private static void sendNotificationToRoki(String deviceToken, String title, String body) throws FirebaseMessagingException {
42         Message message = Message.builder()
43             .putData("title", title)
44             .putData("body", body)
45             .setToken(deviceToken)
46             .build();
47
48         // Invia una notifica push al dispositivo di Roki - Il cellulare in specifico
49         String response = FirebaseMessaging.getInstance().send(message);
50         System.out.println("Messaggio inviato con successo: " + response);
51     }
52
53     /* In seguito l'App Mobile di Roki deve essere configurata per ricevere notifiche
54     push da Firebase, quindi devi prima registrare l'app con Firebase, ottenere un token
55     per il dispositivo e gestire le notifiche in arrivo.
56     */
57 }
58
59

```

Questo esempio, fornisce un programma semplice e teorico che funge da base da cui partire per costruire un sistema di rilevamento di intrusioni basato su IA con notifiche push.

L'implementazione completa, ovviamente richiederà un lavoro significativo e non indifferente per integrare tutte le componenti, configurare i servizi cloud, e assicurare la sicurezza e la privacy dei dati analizzati.

Dopo aver analizzato tutto questo possiamo concludere che il rilevamento delle intrusioni alimentato dalle IA rappresenta un'evoluzione critica nella difesa informatica, trasformando tutto il modo in cui noi e le organizzazioni ci proteggiamo dalle minacce cybernetiche emergenti e vengono così stabiliti nuovi standard di sicurezza.

Conclusioni e ringraziamenti

Con questa tesi si segna il termine del mio viaggio accademico in Ingegneria Informatica.

Desidero esprimere la mia gratitudine al professor Roberto Caldelli per la sua guida indispensabile nel mio lavoro di ricerca.

Un ringraziamento speciale va alla mia famiglia, e in particolare a mia madre, il cui sostegno incondizionato è stato fondamentale.

Un grazie ai miei amici, che hanno giocato un ruolo essenziale offrendomi compagnia e sostegno costanti.

Infine, un ringraziamento particolare va a Gabriele, la cui motivazione è stata decisiva per intraprendere il mio percorso universitario.

Grazie a tutti.

Bibliografia

- [1] Marvin Minsky
- [2] G. Balinggan, 2019. [Online]. Available:
[http://www.differencebetween.net/science/differencebetween-artificialintelligence-and-human-intelligence/..](http://www.differencebetween.net/science/differencebetween-artificialintelligence-and-human-intelligence/)
- [3] Nvidia, «Nvidia Jetson,» [Online]. Available:
<https://www.nvidia.com/itit/autonomous-machines/embedded-systems>.
- [4] Nils J. Nilsson, "Artificial Intelligence: A New Synthesis"
- [5] Stuart Russel and Peter Norvig, "Artificial Intelligence: A Modern Approach"
- [6] Scott Aaronson, "Quantum Computing Since Democritus"
- [7] Negnevitsky, Michael. "Artificial Intelligence Techniques." Editore XYZ, 2022. Capitolo 2.4.
- [8] Robert Morris
- [9] <https://cips.it/autenticazione-vs-autorizzazione-come-comprendere-la-differenza/>
- [10] <https://www.fastweb.it/fastweb-plus/digital-magazine/crittografia-end-to-end-che-cos-e-e-come-funziona/>
- [11] <https://www.cybersecurity360.it/soluzioni-aziendali/crittografia-quantistica-cos-e-come-usarla-per-garantire-massima-protezione-ai-dati-sensibili/>
- [12] <https://www.geopop.it/cosa-sono-gli-attacchi-informatici-e-quali-sono-le-principali-tipologie-e-le-piu-diffuse/>
- [13] <https://www.cybersecurity360.it/soluzioni-aziendali/rischi-di-sicurezza-dei-sistemi-di-intelligenza-artificiale-generativa-limpatto-nel-mondo-cyber/>
- [14] <https://cyberment.it/cyber-attacchi/virus-informatici-ai/>

- [15] <https://www.agendadigitale.eu/cultura-digitale/intelligenza-artificiale-per-la-pubblica-sicurezza-utilizzi-e-rischi-sociali/>
- [16] <https://cyberment.it/cyber-attacchi/virus-informatici-ai/>
- [17] <https://openai.com/blog/chatgpt>
- [18] <https://www.wired.it/article/wormgpt-come-funziona/>
- [19] <https://www.zerounoweb.it/techtarget/searchsecurity/deepfake-phishing-pericolo-ransomware-intelligenza-artificiale/>
- [20] <https://www.zerounoweb.it/techtarget/searchsecurity/deepfake-phishing-pericolo-ransomware-intelligenza-artificiale/>
- [21] <https://www.dirittobancario.it/art/luso-dellintelligenza-artificiale-nella-finanza/>
- [22] <https://www.cybersecurity360.it/nuove-minacce/intelligenza-artificiale-e-machine-learning-nelle-mani-dei-cyber-criminali-rischi-e-contromisure/>
- [23] <https://www.cybersecurity360.it/soluzioni-aziendali/attacchi-sofisticati-difendersi-con-intelligenza-artificiale-machine-learning-e-automazione/>