INTELLIGENZA ARTIFICIALE E SICUREZZA

OPPORTUNITA' RISCHI E RACCOMANDAZIONI

Clusit Community For Security

AI Artificial Intelligence

Sommario

1 Executive Summary

L'intelligenza artificiale (nel seguito citata come IA) è un tema di grande attualità per i cittadini e le organizzazioni. La si vede citata nelle pubblicità di prodotti di consumo, dai cellulari ai televisori e si parla sempre più di digitalizzazione dei processi grazie alle nuove tecnologie, fra le quali l'IA gioca un ruolo di rilievo. Una concreta dimostrazione dell'impatto dell'IA e del suo valore economico è fornita dai dati pubblicati nel settembre 2020 da alcuni media: il mercato italiano dell'intelligenza artificiale vale 240 milioni di euro, +20% rispetto al 2019, di cui l'80% commissionato a livello nazionale e il 20% di export. Più del 60% delle aziende italiane si sta muovendo in questa direzione[1].

Sono numerosi i settori economici ove si registrano esperienze di applicazioni basate sull'IA: marketing e vendite, utility ed energia, banche, finanza, assicurazioni, telecomunicazioni, servizi cloud, industria, logistica, industria automobilistica, amministrazione pubblica, giustizia e avvocatura, e sanità. Significativi esempi di successo sono relativi a strade intelligenti, smart city, smart building, infrastrutture critiche e veicoli autonomi.

L'IA giocherà un ruolo di grande protagonista nello sviluppo dei mercati e nell'innovazione dei processi, generando valore, opportunità e miglioramento dell'efficienza. Imprenditori e manager non potranno ignorarne l'incredibile potenziale.

L'intelligenza artificiale studia da circa 70 anni l'emulazione dell'intelligenza da parte di strumenti informatici software e, in certi casi, hardware. La principale area di ricerca e applicazione dell'intelligenza artificiale è il machine learning (algoritmi che imparano e si adattano in base ai dati che ricevono) e in particolare il suo sottoinsieme relativo alle "reti neurali" (modelli matematici composti da neuroni artificiali) e il "deep learning" (reti neurali di maggiore complessità). Appartengono al mondo dell'IA anche il "riconoscimento vocale", la "robotica avanzata" e i "sistemi esperti", molto in voga qualche decennio fa. Tradurre la ricerca in opportunità di business non è immediato. Imprenditori e manager non sempre possiedono tutte le necessarie competenze matematiche e informatiche, ma devono cercare di coglierne appieno le opportunità attraverso fiuto e fantasia.

L'IA è anche uno strumento per rinforzare la sicurezza informatica nelle organizzazioni (nelle quali convivono applicazioni informatiche di base, sistemi di automazione d'ufficio, soluzioni IoT, OT e di intelligenza artificiale). Due esempi per tutti: il crescente utilizzo dell'IA per la gestione dei SOC (security operation center) e l'automazione dei controlli. Come confermato dagli analisti di settore, il traffico Internet legato al business aumenterà di tre volte nel lasso temporale 2017-2023. Senza strumenti basati sull'IA sarà quindi sempre più difficile per gli analisti di sicurezza poter monitorare efficacemente i volumi, le velocità e le varietà di dati e log generati dai presidi di sicurezza e le informazioni provenienti dall'esterno.

L'IA non è solo un'opportunità. Per esempio, nel Global Risk Report redatto dal World Economic Forum nel 2014 era indicato come un possibile rischio per l'esistenza dell'essere

[1] Milano finanza 19 settembre 2020. Articolo di Antonella Ladisi

umano, al pari di terremoti, tsunami, siccità o di batteri resistenti agli antibiotici. Quali sono dunque i principali rischi che presenta l'IA e quali le azioni da mettere in campo per limitarli? Quali sono il contesto normativo e le implicazioni di controllo interno e di sicurezza da considerare quando si tratta di IA? Quali sono le ricadute in tema di protezione dei dati personali e GDPR?

Al proposito si evidenzia l'incapacità di illustrare compiutamente i modelli generati dagli algoritmi, poiché possono cambiare autonomamente anche in modo inatteso e imprevedibile. Questo impone considerazioni etiche quando il loro uso può avere impatti sulla vita, la dignità e la sicurezza delle persone.

Vi sono due classi di minacce: quelle ai sistemi di IA in quanto tali (errori di programmazione, errori di configurazione, inaffidabilità dei dati di input) e quelle ove il malintenzionato utilizza l'IA per condurre attacchi, beneficiando di automatizzazione e autoapprendimento (attività di raccolta dati attiva e passiva, offline e online; pianificazione; phishing; deepfake, ecc.).

Compliance e rischi connessi all'IA richiedono da parte del management di un'organizzazione una particolare attenzione nello svolgimento delle classiche prassi di controllo interno, con particolare riferimento all'audit, al project e al risk management, soprattutto per assicurare la trasparenza delle soluzioni di IA fin dalla progettazione, con applicazione dei principi di privacy e security by design.

Le organizzazioni sono pronte a cogliere le opportunità dell'IA? Allo stato attuale è ragionevole considerare l'IA una novità "disruptive" per due ragioni fondamentali: da un lato, la difficoltà dei vertici e dei manager nel comprenderne l'impatto potenziale sui processi; dall'altro lato, la preparazione delle persone dedicate alla progettazione e realizzazione di soluzioni di IA in area tecnologica e organizzativa. Come superare tale situazione che costituisce una barriera alla possibilità di cogliere le innegabili opportunità dell'IA? Potrebbero essere considerate due alternative: percorrere strade già collaudate in passato per la gestione dell'innovazione ricorrendo alla leva formativa e promuovendo un progressivo cambiamento nei profili di riferimento manageriali e tecnologici, oppure cercare nuovi approcci di non semplice identificazione.

Probabilmente la soluzione ideale da perseguire potrebbe essere un mix delle alternative citate, come di fatto è avvenuto sul mercato a fronte di cambiamenti rilevanti come l'avvento dei sistemi informatici real-time nel decennio 1965-75, o l'avvento di Internet negli anni novanta. In questo scenario acquistano grande rilevanza la gestione delle risorse umane e il sistema formativo. Queste due dimensioni assumono una dimensione strategica in quanto sono i capisaldi per una pianificazione consapevole del patrimonio di competenze e di profili professionali che ormai le organizzazioni non possono più trascurare se vorranno cogliere le innegabili opportunità dell'intelligenza artificiale.

La Community for security del Clusit si augura che questa pubblicazione possa fornire ai lettori un utile quadro d'insieme di una realtà, come l'intelligenza artificiale, che li accompagnerà sempre più nella vita personale, sociale e lavorativa.

- PRIMA PARTE: INTRODUZIONE E PRESENTAZIONE

2 Per chi abbiamo scritto questo libro

L'IA è una tematica che, potenzialmente, può interessare tutte le organizzazioni, di qualunque settore economico, dalle grandi organizzazioni alle PMI, dalle aziende private alla PA. Nell'executive summary sono riportati cenni sui principali settori dove sono stati già realizzati, o sono in corso di realizzazione, progetti per l'uso dell'IA, in particolare nelle grandi aziende, per conseguire benefici gestionali. La pubblicazione è pertanto destinata a tutti i soggetti che operano sul mercato, dalle grandi aziende che, per svariati motivi, non hanno ancora sperimentato soluzioni di IA, alle PMI che potrebbero considerare l'opzione IA per migliorare le soluzioni interne (vendite, produzione, logistica e - perché no? - innovazione). Le organizzazioni che già adottano soluzioni basate sull'IA potrebbero beneficiare di questa pubblicazione per approfondirne le opportunità, i rischi e le relative contromisure. Anche la PA, in coerenza con la spinta alla digitalizzazione, a livello europeo e nazionale, può trovare validi riferimenti e spunti utili a stimolare l'efficienza interna e di servizio verso i cittadini e gli operatori economici privati.

Nei numerosi casi di adozione di soluzioni di IA trattati nella pubblicazione, i lettori potranno trovare spunti concreti applicabili in molti contesti, che potrebbero stimolare riflessioni su come promuovere l'introduzione dell'IA all'interno dell'organizzazione di appartenenza, considerando le diverse sfaccettature che questa scelta richiede di valutare. Gli aspetti da considerare al riguardo vanno dalla individuazione dei processi interni "candidabili" per la adozione della IA, alle modalità più opportune per promuovere la costituzione dei gruppi di progetto, alle esigenze informative e formative necessarie, all'individuazione di eventuali vincoli legali, organizzativi, culturali, tecnologici e, soprattutto, ai fattori "abilitanti" che potrebbero favorire l'introduzione nelle organizzazioni dell'IA e che, in sintesi, dovrebbero sempre essere ricondotti allo scenario di mercato nazionale e internazionale e al sistema competitivo nel quale tutte le aziende, anche le PMI, si trovano e si troveranno sempre più ad operare.

L'IA si può quindi applicare ad un'ampia gamma di settori industriali e di applicazioni eterogenee e richiede di considerare alcuni aspetti come la normativa vigente, la sicurezza e i rischi in genere ad essa collegati. Questa premessa fa intuire che non esiste una figura professionale primaria a cui la pubblicazione è dedicata; certo, il Clusit si occupa di sicurezza informatica, e ciò fa quindi supporre che i principali destinatari dell'opera siano i CISO delle organizzazioni, ma non è così.

Lo sforzo fatto dagli autori ha invece permesso di redigere un'opera che non si rivolge solo a specifici interlocutori ma che, anzi, è di utilità per un ampio spettro di professionalità. Il *responsabile della sicurezza* troverà tutti quegli elementi che gli permetteranno di valutare i rischi e predisporre contromisure. Il *project manager* capirà come gestire meglio progetti di IA. Il *data scientist* troverà particolarmente utile leggere il capitolo sugli algoritmi.

Gli *analisti IT* e gli *esperti di organizzazione* potranno farsi un'idea concreta delle potenziali implicazioni dell'IA con il sistema informativo e con il modello organizzativo aziendale.

I *responsabili* delle aree di ricerca e sviluppo e i *manager* in generale potranno appassionarsi con i tanti esempi di applicazioni reali dell'IA, che saranno di ispirazione per la propria realtà organizzativa e che potranno ispirare anche i *system integrator*; le *figure legali* e i *DPO* troveranno interessanti gli spunti relativi alla relazioni tra IA e normativa vigente, con particolare riferimento alla privacy, al diritto d'autore e alla responsabilità civile e penale.

Tutti coloro che vogliono ottimizzare e rendere più efficienti i propri processi aziendali, i *responsabili della produzione* o *ingegneri di processo*, gli specialisti di IoT e OT e di digitalizzazione che vogliono innovare e progettare nuovi prodotti e i *progettisti di dispositivi e di macchine* troveranno questa lettura interessante ed istruttiva.

Quindi, se pensate che questa opera non sia per voi, leggetela e vi ricrederete.

3 Obiettivi e sintesi

La pubblicazione è stata concepita per favorire un progressivo avvicinamento del lettore al mondo dell'IA, in un percorso che inizia da tematiche di natura concettuale e metodologica e si sviluppa con concreti esempi di applicazione dell'IA nelle organizzazioni, per concludersi con valutazioni relative ad aspetti specifici (ad esempio compliance, etica, rischi e sicurezza) e con interviste a persone con esperienze in progetti di IA.

La **prima parte** "Introduzione e presentazione" è centrata sull'individuazione dei potenziali lettori della pubblicazione: le organizzazioni in generale e le singole persone coinvolte in relazione al loro ruolo nelle organizzazioni. La **seconda parte** "Che cos'è l'intelligenza artificiale" delinea le principali componenti dell'IA. Sono riportate le più diffuse definizioni dell'IA, approfondendone gli ambiti di applicazione, le architetture hardware e software, le diverse tipologie di algoritmi utilizzabili, le diverse modalità di apprendimento, i possibili approcci allo sviluppo dell'IA e, infine, le principali entità da considerare nei progetti di IA: dati, fasi progettuali, integrazione dell'IA nei processi aziendali. Definiti i potenziali lettori e individuati i "mattoni" da utilizzare per la costruzione di soluzioni di IA, la pubblicazione, nella **terza parte**, elenca alcuni possibili "Usi dell'IA" per consentire al lettore una comprensione pratica dell'IA. L'IA non è dominio esclusivo degli specialisti della materia (negli anni Ottanta si identificavano come "ingegneri della conoscenza"), ma richiede una condivisione anche con i ruoli aziendali dedicati alla gestione del business e dei processi aziendali. Ci si focalizza infatti sul valore delle soluzioni di IA con la descrizione di una quarantina di progetti realizzati in una ventina di settori, tutti di grande interesse e attualità, cui si rimanda per una presa di conoscenza adeguata e per cogliere eventuali spunti di applicabilità nei contesti in cui operano i lettori.

Nella **quarta parte** "Normativa ed Etica", il focus si sposta sugli aspetti di compliance che riguardano in particolare privacy, brevetti, responsabilità in ambito civilistico e penale, nonché la dimensione strategica per il futuro dell'IA. In particolare è approfondito il principio di trasparenza nelle soluzioni di IA, che ha riflessi fondamentali per la compliance nei trattamenti automatizzati di dati personali e, soprattutto, per la dimensione etica verso la collettività, le singole persone e i portatori di interesse di un'organizzazione. La **quinta parte** è centrata sui rischi dell'IA, che interessano più dimensioni: i rischi per le persone, i rischi connessi all'integrazione di tecnologie che convivono con le soluzioni di IA (ad esempio nelle fasi di raccolta dei dati), le minacce ai sistemi di IA, e anche le minacce derivanti da attacchi che possono essere condotti con il supporto dell'IA. Nella **sesta parte** "Contromisure**"** ci si addentra in una dimensione familiare per la Community for Security di Clusit, facendo particolare riferimento alla sicurezza logica, infrastrutturale, fisica e alla continuità operativa delle soluzioni di IA, con approfondimenti sulle problematiche assicurative e sul possibile contributo dell'IA per la prevenzione delle frodi, per la sicurezza delle applicazioni IOT e OT e per la sicurezza nazionale.

Nella **settima parte** "Conclusioni" si parla del futuro dell'IA e, successivamente, riferimenti a standard e best practice relative all'IA.

In tutto il testo sono riportate alcune interviste svolte con persone aventi esperienze qualificanti in progetti di IA.

La consultazione della pubblicazione è un impegno non trascurabile per il lettore interessato alla tematica IA, considerata da molti angoli visuali. Pur essendo difficile immedesimarsi nel lettore, è ragionevole suggerire un percorso di lettura diversificato: per tutti partire dall'executive summary e dall'introduzione, e, in relazione all'interesse del lettore, orientarsi sulle parti che trattano di soluzioni ed esperienze di IA, oppure approfondire gli aspetti specifici riportati nelle altre parti del documento.

Buona lettura.

4 Pubblicazioni della Community e licenza

Questa è la dodicesima pubblicazione della Clusit Community for Security, senza contare il blog dedicato al Regolamento europeo sulla protezione dei dati personali, con alcune centinaia di post in italiano e in inglese[2].

Gli autori svolgono il lavoro di preparazione tramite un confronto multidisciplinare e multisettoriale. Li motiva la consapevolezza del grande bisogno di sicurezza e compliance delle aziende italiane e un forte senso di responsabilità verso la nostra società.

La Community opera dal 12 settembre 2007 e permette la collaborazione di alcune centinaia di professionisti che operano negli ambiti della sicurezza, dell'audit, della conformità, dell'ethical hacking, della consulenza, dell'integrazione dei sistemi e delle certificazioni basate su norme internazionali. Partecipano alla Community i responsabili della sicurezza del mondo della domanda e dell'offerta di servizi e tecnologie correlati a questi ambiti. Il lavoro verte su molteplici aspetti: culturale, organizzativo e tecnologico. La Community riceve il sostegno di prestigiose associazioni professionali e industriali che collaborano attivamente tramite i loro membri, come per esempio: ABI Lab, ACFE, AIEA, AISIS, ANDIP, ANORC, APIHM, AUSED, BCI Italy Chapter, CSA Italy, ISC2.

Le pubblicazioni fatte finora sono liberamente scaricabili dal sito http://c4s.clusit.it e sono:

- ROSI - Return on Security Investments: un approccio pratico. Come ottenere Commitment sulla Security;
- Fascicolo Sanitario Elettronico: il ruolo della tecnologia nella tutela della privacy e della sicurezza;
- Privacy nel Cloud: Le sfide della tecnologia e la tutela dei dati personali per un'organizzazione italiana;
- Mobile e Privacy: Adempimenti formali e misure di sicurezza per la conformità dei trattamenti di dati personali in ambito aziendale;
- La sicurezza nei Social Media: guida all'utilizzo sicuro dei Social Media per le aziende del Made in Italy;
- I primi 100 giorni del responsabile della Sicurezza delle Informazioni: Come affrontare il problema della Sicurezza informatica per gradi;
- Le frodi nella rete: il duplice ruolo dell'ICT;
- Mobile Enterprise: sicurezza in movimento;
- SOC e Continuous Monitoring faccia a faccia con la Cybersecurity;
- Consapevolmente Cloud. Guida per l'azienda che deve affrontare l'innovazione con le idee chiare;
- IoT Security e Compliance. Gestire la complessità e i rischi.

[2] http://europrivacy.info è stato recentemente ceduto gratuitamente a Mandat International, organizzazione non governativa internazionale con sede a Ginevra, che propone uno schema di certificazione privacy sviluppato nell'ambito del programma di ricerca europeo Horizon 2020. Europrivacy è gestito da European Centre for Certification and Privacy.

Queste pubblicazioni sono il frutto del lavoro di almeno 50 persone e consistono in 100-200 pagine di materiale. Consci che tutto è migliorabile, e nel pieno spirito della Community, le rendiamo disponibili con una licenza “Creative Common, Attribuzione e Condividi nello stesso modo” (https://creativecommons.org/licenses/by-sa/4.0/). La licenza permette a chiunque di usare il nostro prodotto per crearne una sua evoluzione a condizione che citi gli autori originali riportando la nostra URL http://c4s.clusit.it e utilizzi a sua volta lo stesso tipo di licenza.

- SECONDA PARTE: CHE COS'È L'INTELLIGENZA ARTIFICIALE

5 Le definizioni dell'intelligenza artificiale

5.1 Gli ambiti dell'intelligenza artificiale

L'intelligenza artificiale è una disciplina che da circa settant'anni studia l'emulazione dell'intelligenza da parte delle macchine intese come software e – in certi casi – hardware. Si tratta di un campo molto vasto, data anche l'intrinseca difficoltà che si ha nell'arrivare a una definizione unanime e universale, come il lettore potrà approfondire nel paragrafo successivo.

Storicamente, la ricerca nell'ambito dell'intelligenza artificiale nasce dalla difficoltà di risolvere problemi usando algoritmi deterministici, ossia nei casi in cui è molto difficile descrivere in maniera precisa il compito dell'algoritmo. Si consideri ad esempio il problema del riconoscimento delle cifre da 0 a 9 scritte a mano (in figura è presentato un esempio del dataset MNIST, creato dallo statunitense NIST, che si può usare a questo scopo) o del riconoscimento di un volto: in entrambi i casi, è particolarmente arduo tradurre in un algoritmo deterministico il processo di riconoscimento messo in atto dal cervello umano.

Figura 1 - Esempi di cifre scritte a mano, tratti dal dataset MNIST[3]

Le aree di ricerca e applicazione dell'intelligenza artificiale sono molteplici e persino compilarne un elenco o tracciarne un diagramma porta a risultati non accettati da tutti poiché, man mano che la disciplina si sviluppa, nuovi metodi si intrecciano e si evolvono arrivando a collegare e sovrapporre settori un tempo distinti e distanti.

[3] Fonte: https://commons.wikimedia.org/w/index.php?curid=64810040.

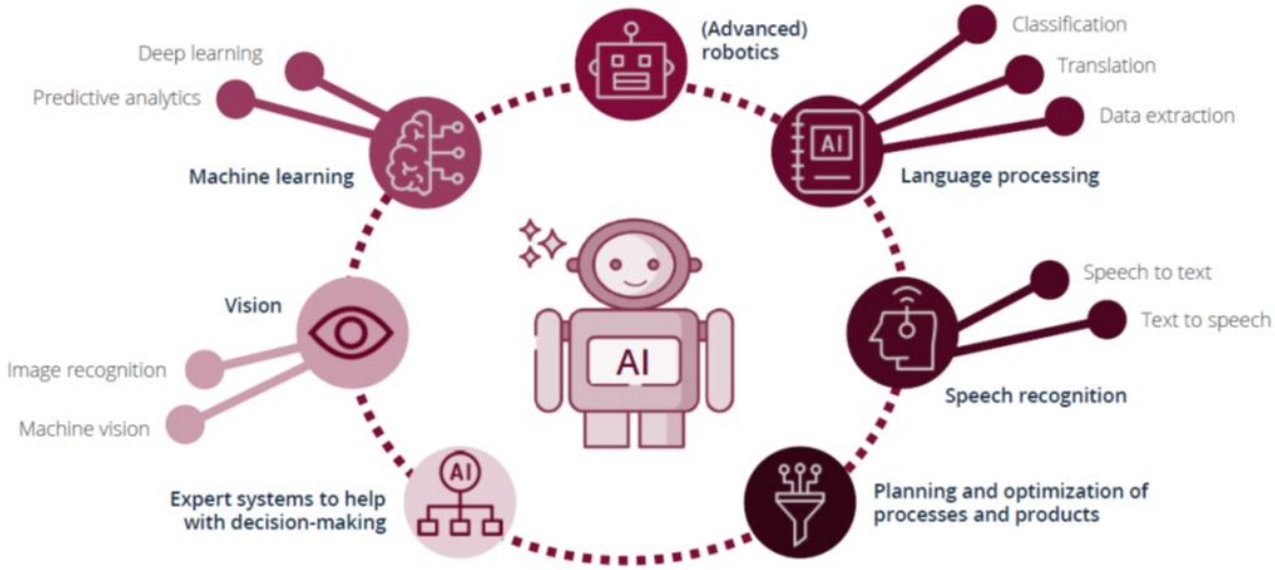


Figura 2 - Gli ambiti di ricerca dell'intelligenza artificiale[4]

Gli schemi più accreditati includono, nella vasta disciplina dell'intelligenza artificiale, le aree del machine learning, della percezione, dell'elaborazione del linguaggio naturale e della robotica.

Machine learning: in quest'area si studiano algoritmi che imparano e si adattano in base ai dati che ricevono. Essi vengono addestrati, solitamente con insiemi di dati (**dataset**), come illustrato più estesamente al paragrafo 7.2. Una volta messi in produzione, questi sistemi si aggiornano attraverso i dati che incontrano. Per fare un esempio, un filtro anti-spam viene addestrato dai ricercatori a riconoscere messaggi di posta indesiderata attraverso un dataset contenente posta normale e spam; tuttavia, una volta attivo nella casella di posta dell'utente, esso continua ad aggiornare i propri parametri a seconda di ciò che l'utente etichetta come spam oppure posta normale.

Il machine learning è suddiviso in diverse **tipologie** (vedere anche il paragrafo 7.2), le cui principali sono:

- **apprendimento supervisionato**, che a sua volta si divide in **classificazione** (in cui è necessario etichettare i dati a mano per insegnare all'algoritmo a classificarli) e **analisi della regressione** (che adotta tecniche statistiche per stimare eventuali interazioni fra le variabili, arrivando a prevedere trend e pattern);
- **apprendimento non supervisionato**, in cui è l'algoritmo stesso che crea le etichette a seconda di ciò che trova nei dati, raggruppandoli quindi secondo "cluster" non necessariamente previsti dagli sviluppatori;
- **apprendimento per rinforzo**, in cui gli algoritmi prendono decisioni in base a obiettivi prefissati, cercando di massimizzare i risultati.

Per fare un esempio, un'azienda che svolge attività di data mining per studiare i comportamenti dei propri clienti – arrivando così a previsioni di acquisto o di utilizzo dei servizi – utilizza il machine learning supervisionato, in particolare la regressione.

[4] Fonte: Thomson Reuters/Statista.

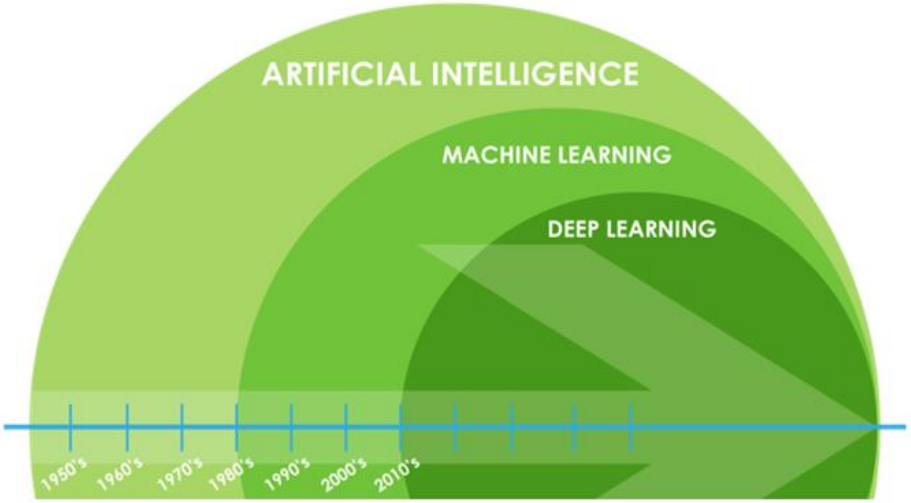


Figura 3 - Le relazioni fra deep learning, machine learning e IA, con linea temporale di sviluppo[5]

Numerosi sono gli utilizzi del machine learning, in particolare delle **reti neurali** e del **deep learning**, che costituiscono, a tutti gli effetti, un sottoinsieme del machine learning.

I metodi per progettare un algoritmo di machine learning sono numerosi e includono gli **alberi decisionali** (p.e. il tipico diagramma a flusso dove si rappresentano le possibili variabili), le **macchine a vettori di supporto (support vector machines** o **SVM,** descritte nel paragrafo 7.4.3) e le **reti bayesiane** (vedere paragrafo 7.4.3); tuttavia, in questi ultimi anni, le **reti neurali** sono state di gran lunga la metodologia più utilizzata. Una rete neurale è composta da nodi, chiamati anche neuroni artificiali, disposti in strati (layer) con uno strato di input, almeno uno strato interno (o "nascosto") e uno strato di output.

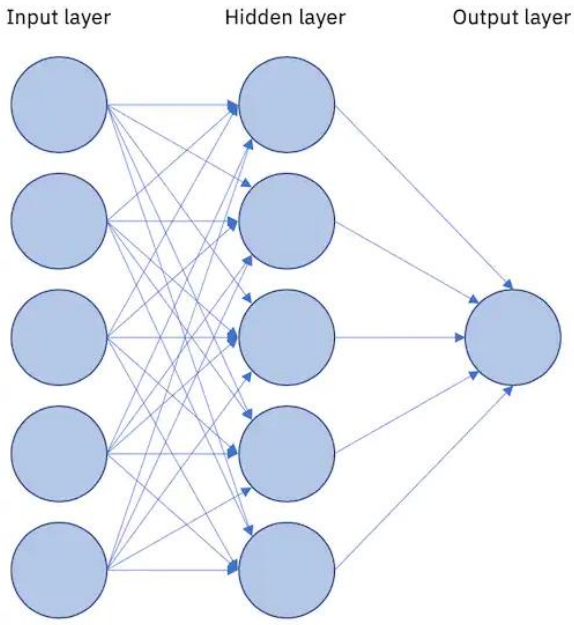


Figura 4 - Una rete neurale con tre strati[6]

[5] Fonte: Buzzrobot.
[6] IBM Cloud.

Le reti neurali sono tanto più potenti quanto più si aumenta il numero di strati nascosti. Nelle reti più grandi si arriva a milioni di neuroni artificiali e decine di miliardi di connessioni (a cui sono associati dei “pesi”, che sono i parametri variabili del modello e vengono tarati durante la fase di addestramento). In questo caso si parla di **reti neurali profonde**, o deep neural networks (DNN), usate per attività di **deep learning**. È possibile trovare maggiori informazioni nel paragrafo 7.4.1 sull'approccio connessionista.

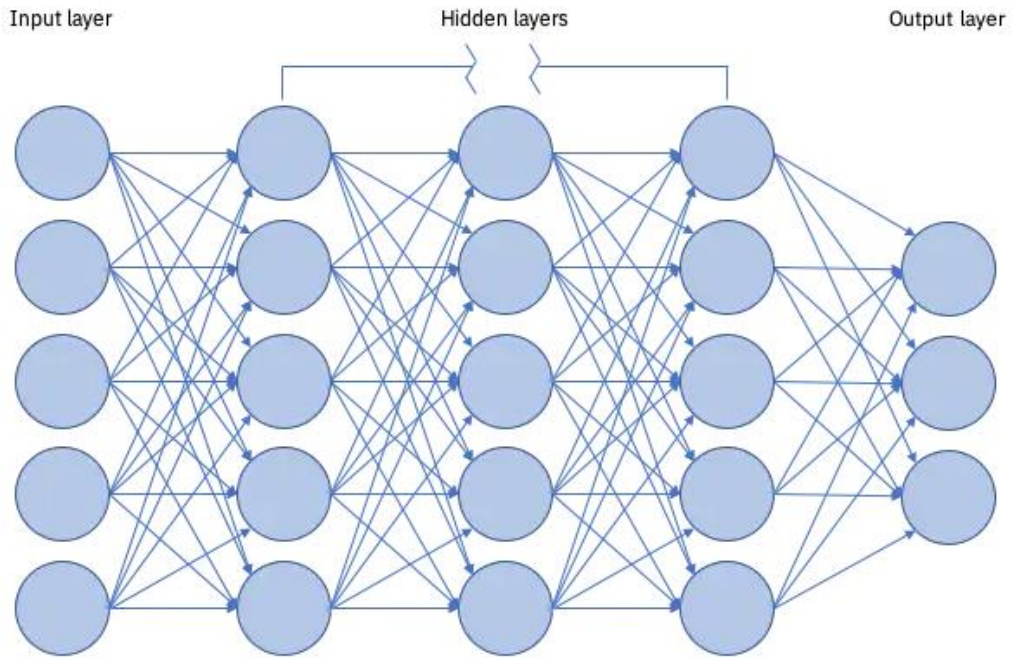


Figura 5 - Una rete neurale “profonda” con più di tre strati[7]

Percezione: l’ambito della percezione racchiude principalmente due aree sul piano informatico molto diverse fra loro: la **visione artificiale** e il **riconoscimento vocale**.

La visione artificiale viene usata soprattutto per esaminare e classificare oggetti (ad esempio, riconoscere il contenuto di un'immagine: animali, persone, marchi commerciali, e altri oggetti), consentire a un agente fisico di orientarsi (esempi sono un robot che deve seguire un percorso e il riconoscimento della segnaletica stradale per le auto a guida autonoma), individuare le persone e identificare i tratti somatici del volto (in questo caso si parla di riconoscimento facciale). Una delle applicazioni più promettenti, per fare un esempio fra tanti, si ha in ambito medico, in cui la visione artificiale viene impiegata per offrire un’indicazione probabilistica delle patologie dei pazienti, partendo da immagini strumentali come TAC, ecografie e risonanze magnetiche.

Il riconoscimento vocale, invece, individua il linguaggio da un input audio e lo traduce in testo. I modelli di riconoscimento vocale già oggi disponibili[8], oltre a riconoscere quello che viene detto, possono essere addestrati per identificare una persona, partendo dalle caratteristiche

[7] IBM Cloud.
[8] https://www.sciencedirect.com/topics/computer-science/speaker-identification.

della sua voce, isolare una o più voci in mezzo a tante altre e, addirittura, suggerire la presenza di eventuali patologie rilevando microelementi nel tono, nel ritmo e nella frequenza della voce[9].

I risultati migliori in ambito di percezione – a testimonianza di come i vari ambiti dell'intelligenza artificiale si intersecano fra loro – si raggiungono quando i dati sono elaborati dalle reti neurali attraverso algoritmi di deep learning.

Elaborazione del linguaggio naturale: è noto anche come *natural language processing* (NLP) e si occupa della comprensione e della produzione del linguaggio umano. Si tratta di un ambito particolarmente significativo per la reale comprensione dell'IA da parte delle persone comuni: mentre la maggior parte delle attività di IA sono di difficile comprensione dalla persona media, l'interazione diretta fra una persona e una macchina, che comprende e risponde in maniera pertinente e con un linguaggio naturale, è immediatamente evidente.

L'NLP è un ambito in cui si lavora da oltre settant'anni; inizialmente i modelli erano basati su regole complesse inserite manualmente dai ricercatori, come ad esempio le ontologie concettuali. A partire dagli anni Novanta si iniziò a usare il machine learning per addestrare modelli di traduzione e, negli ultimi anni, con l'arrivo del deep learning – in particolare un modello del 2017 noto come "Transformer" – si sono raggiunti risultati di comprensione e di produzione di testi senza precedenti.

Robotica (avanzata): la robotica è per molti aspetti una disciplina a sé, alla quale non servirebbe l'intelligenza artificiale per esistere. La robotica avanzata, tuttavia, integra molte delle tecnologie IA che abbiamo visto finora (visione computerizzata, riconoscimento ed elaborazione del linguaggio, elementi di machine learning, ecc.) per arrivare a decisioni e risultati in autonomia. Anche le auto a guida autonoma rientrano appieno nella robotica avanzata, essendo robot che usano numerose tecniche di intelligenza artificiale per raggiungere i propri obiettivi.

Elemento caratterizzante della robotica sono gli agenti fisici che abitano lo stesso mondo in cui viviamo. Al contrario dei software IA, per loro natura immateriali e senza ubicazione fisica ben definita, un robot occupa spazio, si muove, interagisce fisicamente con gli esseri umani, potendo – per queste caratteristiche – causare incidenti fisici anche gravi.

Tutti i robot possono contenere al loro interno software che impartiscono loro istruzioni su cosa fare e come farlo. Tali programmi sono essenzialmente di tre tipi:

- i **robot a controllo remoto** contengono istruzioni che vengono generalmente attivate da esseri umani (p.e. un ascensore);
- i **robot con intelligenza artificiale** possono decidere autonomamente come comportarsi in base alle circostanze, con una serie di opzioni che variano da decisioni molto semplici (p.e. i robot-giocattolo che, se incontrano un ostacolo, si fermano) a decisioni molto avanzate (p.e. le auto a guida autonoma, che mantengono la corsia e la velocità adeguata alla segnaletica stradale);

[9] https://www.fanaticalfuturist.com/2017/02/artificial-intelligence-diagnoses-disease-by-listening-to-your-voice/.

- i **robot con programmazione ibrida** posseggono elementi IA ma possono essere controllati da remoto in caso di necessità, come molti robot industriali nelle catene di montaggio.

Altri ambiti di impiego dell'IA: vi sono molti altri ambiti e tecniche che, in un modo o nell'altro, afferiscono almeno parzialmente alla disciplina. Un esempio sono i **sistemi esperti,** molto in voga qualche decennio fa e oggi caduti in disuso grazie alla rivoluzione del deep learning. I sistemi esperti riproducono le prestazioni di una persona esperta di un determinato argomento e riescono a dedurre nuovi fatti da quelli già noti. I sistemi esperti dispongono sempre di una base di conoscenza iniziale e quindi non appartengono alla categoria del machine learning.

Il lettore, all'interno del libro, avrà la possibilità di approfondire molte di queste tecnologie nei capitoli specifici, sempre tenendo presente che tutto quello che **fa apparire** una macchina come dotata di intelligenza può considerarsi a pieno diritto parte del grande e ospitale mondo dell'intelligenza artificiale.

5.2 Definire l'intelligenza artificiale

"L'intelligenza artificiale (IA) è la capacità delle macchine o dei computer di eseguire compiti che vanno oltre la loro programmazione, come l'apprendimento e la soluzione di problemi. Il termine intelligenza artificiale *è stato applicato in molti modi diversi, ma il suo uso più comune si riferisce a programmi per computer che hanno imparato comportamenti di tipo umano attraverso una vasta esperienza. Ciò include elementi come il riconoscimento vocale, il* pattern matching*, l'analisi delle immagini e molto altro ancora. Affinché una macchina diventi intelligente deve imparare dalle esperienze passate e avere la capacità di ragionare su nuove situazioni. Queste due abilità sono strettamente correlate e sono necessarie l'una all'altra. La macchina deve inoltre dimostrare capacità di adattamento in risposta a circostanze mutevoli. Ciò è particolarmente importante se il suo scopo è quello di interagire con l'uomo o con la società in generale: un robot che non possa cambiare il suo comportamento quando necessario non sarebbe considerato intelligente da nessuna definizione del termine".*

Il paragrafo che avete appena letto è una definizione di intelligenza artificiale prodotta dalla stessa IA[10] nella forma di un modello di linguaggio autoregressivo denominato GPT-3[11]. Abbiamo dato la parola all'intelligenza artificiale affinché avesse anche lei la possibilità di dire la sua su un argomento molto dibattuto. Questa infatti è solo una fra le tantissime definizioni, creata da una rete neurale di deep learning che ha potuto attingere alla vasta letteratura (scientifica e non) dagli anni Cinquanta a oggi. Nonostante ciò, anche questa risente di alcuni difetti ed "escamotage" che tradiscono una mancanza di elementi di base: affermare ad esempio che l'IA mostra comportamenti "di tipo umano" si affida a un sottinteso tutto da definire, una scorciatoia già usata da molti importanti pensatori umani.

[10] https://twitter.com/LucaSambucci/status/1297471767536119808.
[11] https://arxiv.org/abs/2005.14165.

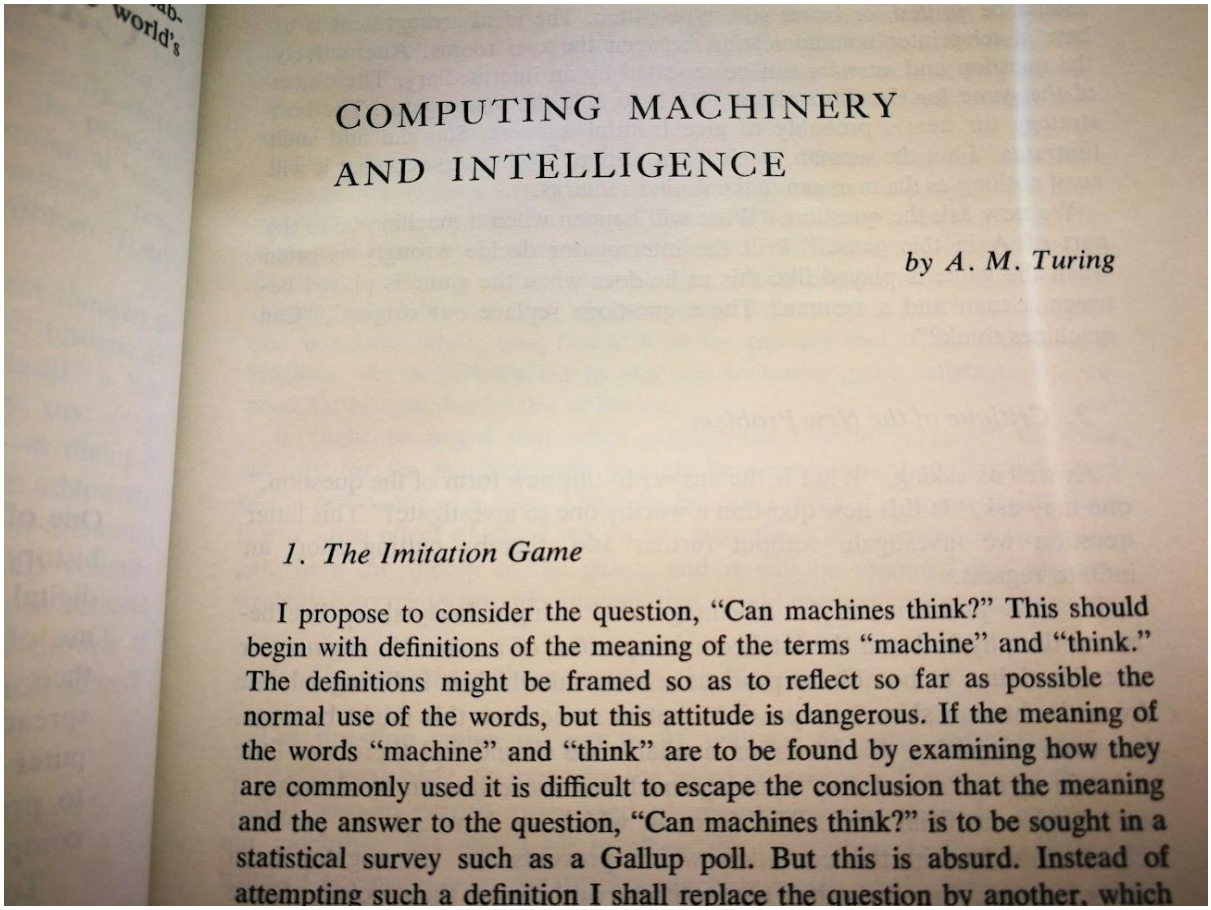
Perché è così difficile definire che cos'è l'intelligenza artificiale? Questa complicazione - che ci portiamo appresso da decenni - ha favorito una certa confusione, che induce molte persone a scambiare la disciplina con i suoi elementi costituenti. L'intelligenza artificiale non è solo *machine learning* (ML), che pure occupa un posto di primo piano nel settore, così come il *machine learning* non inizia e finisce con il *deep learning*, anche se quest'ultima tecnica ha monopolizzato gran parte delle attività di ML. L'IA è composta anche da sistemi esperti, grafi della conoscenza, motori di regole che costituiscono quella che spesso viene definita affettuosamente GOFAI - good old-fashioned artificial intelligence (cara vecchia intelligenza artificiale) - o più seriamente "IA simbolica", dove per "simbolico" si intende una rappresentazione della conoscenza intellegibile agli umani, ottenuta manipolando simboli e "frasi" (anziché numeri), come avviene nell'elaborazione del nostro linguaggio parlato e scritto. La GOFAI, anche se proveniente dal passato dell'IA e nonostante il nome, non è per nulla obsoleta, poiché in diversi casi offre quel contesto che i modelli basati su machine learning ancora non riescono adeguatamente a fornire.

Tornando alle definizioni ritroviamo la difficoltà di articolare cosa voglia dire essere intelligente. Un caso esemplare è la famosa definizione di Marvin Minsky, uno dei pilastri della ricerca nell'ambito della IA, che nel 1968 definì l'intelligenza artificiale come "*la scienza del creare macchine che fanno cose che richiederebbero intelligenza se fatte dall'uomo*"[12], riprendendo quasi parola per parola la definizione pare originata da John McCarthy - vero e proprio padre dell'IA: "*il problema dell'intelligenza artificiale consiste nel creare una macchina che si comporti in modi che sarebbero considerati intelligenti se un umano si comportasse nello stesso modo*"[13]. Quest'ultima definizione fu inserita in una proposta del 1955, preliminare al famoso workshop del 1956 presso il Dartmouth College che diede formalmente inizio alla disciplina.

Queste definizioni saltano a piè pari il problema di definire cosa sia l'intelligenza. Forse anche per questo lo stesso McCarthy, oltre cinquant'anni dopo, nel 2007 (il 12 novembre alle 2 e 05 del mattino, come ha scrupolosamente registrato nel documento), ha pubblicato l'articolo "What is Artificial Intelligence?" che rappresenta una sorta di FAQ sulla natura dell'intelligenza artificiale[14] e che prende di petto proprio questo problema. "*Il problema* - scrive McCarthy - *è che ancora non possiamo caratterizzare in generale quali tipologie di procedure computazionali vogliamo chiamare intelligenti. Noi comprendiamo alcuni dei meccanismi dell'intelligenza, ma non altri*".

Lo stesso Alan Turing si pose la questione su cosa volesse dire "pensare" già nel 1950, nel saggio "Computing machinery and intelligence", considerato uno dei lavori che più hanno influenzato la nascita dell'intelligenza artificiale.

[12] Marvin L. Minsky (1968) in Stonier, T., 1992. Beyond Information. London: Springer, pp.107-133.
[13] http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html.
[14] http://jmc.stanford.edu/articles/whatisai/whatisai.pdf.



COMPUTING MACHINERY
AND INTELLIGENCE

by A. M. Turing

1. The Imitation Game

I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which

Figura 6 - Il saggio "Computing machinery and intelligence" del 1950, in una pubblicazione del 1963[15]

La difficoltà che abbiamo nel definire l'IA è proprio questa: non siamo ancora pienamente padroni di una definizione di intelligenza che trascenda l'origine dell'agente (umano, animale o meccanico) e che metta d'accordo tutti. Senza tale prerequisito non riusciremo a definire con soddisfazione cos'è e cosa non è intelligenza artificiale.

La maggior parte delle definizioni sposta dunque il problema su cosa noi esseri umani percepiamo come intelligente. Del resto anche il famoso "test di Turing"[16] si basa su un criterio simile che, come spiegano Luigia Carlucci Aiello - che ha contribuito a far approdare la disciplina in Italia dopo aver lavorato proprio con McCarthy negli Stati Uniti - e Marta Cialdea Mayer, "*è di tipo puramente comportamentale, così che nel giudicare l'intelligenza di un sistema artificiale non si debbano fare considerazioni su processi mentali o attributi psichici, né tanto meno convenire su una definizione precisa del concetto di intelligenza*"[17].

Ma anche la strada intrapresa da Turing per semplificare il dilemma non convince tutti. Jerry Kaplan, in una recente pubblicazione[18], ha sostenuto che la quantificazione mono-dimensionale che molti di noi hanno dell'intelligenza umana rappresenta un approccio inadeguato per definire l'intelligenza artificiale, capace di calcoli inimmaginabili per un cervello

[15] Il saggio "Computing machinery and intelligence" di Alan Turing del 1950, qui in una pubblicazione del 1963. Fonte: scansione di un volume degli autori di questo libro.
[16] Turing, A.M. *In Mind*, Volume LIX, Issue 236, pp. 433-460. October 1950.
[17] Carlucci Aiello, L. e Cialdea Mayer, M. *Invito all'intelligenza artificiale*. Franco Angeli, 1995.
[18] Kaplan, J.. *Artificial Intelligence*. New York: Oxford University Press, 2016.

umano ma allo stesso tempo incapace di generalizzare con successo da un contesto all'altro. Insomma, può essere considerata intelligente una macchina che elabora milioni di istruzioni al secondo, risolvendo magari un problema al quale non saremmo mai riusciti ad arrivare da soli, ma allo stesso tempo non è in grado di versarci una tazzina di caffè?

O è forse il paragone con l'intelligenza umana a essere sbagliato? Un aeroplano sa volare, come faceva notare Paul Armer nel 1960[19], senza per questo essere in grado di aggrapparsi ai rami come farebbe un uccello. Cercare quindi di antropomorfizzare le macchine, attribuendo loro la qualifica di "intelligenti" solo quando le loro attività sono simili a quelle di un essere umano, è un esercizio che molti considerano sbagliato e per certi versi limitante.

È possibile tuttavia isolare alcuni componenti comuni alla gran parte delle definizioni, fra i quali spicca senz'altro la capacità da parte della macchina di **imparare**. Alan Turing nel 1947, forse in una delle primissime lezioni sull'argomento[20], disse: "*Quello che vogliamo è una macchina che impari dall'esperienza*" e "*la possibilità di consentire alla macchina di alterare le proprie istruzioni offre il meccanismo per realizzare questo*". Questo aprì in pratica le porte al concetto di *machine learning* (ritroveremo molte delle sue idee nello studio "Intelligent machinery" del 1948, purtroppo pubblicato postumo e oggi disponibile online grazie al Turing Archive[21]).

Un altro componente fondamentale che contraddistingue l'intelligenza artificiale è quello di **risolvere problemi**. Uno dei padri dell'intelligenza artificiale italiana, Marco Somalvico, nel 1987 coniugava perfettamente questa caratteristica assieme al concetto di *learning machine* scrivendo: "*si vuole mettere in grado l'elaboratore, non solo di essere un buon esecutore di algoritmi forniti dall'uomo, ma di costruire esso stesso, in modo del tutto automatico, gli algoritmi necessari per risolvere problemi, che l'utente può semplicemente definire, lasciando all'elaboratore la responsabilità della loro soluzione*"[22].

Quindi imparare dalle proprie esperienze e usare queste nozioni - attraverso ragionamenti e processi decisionali - per arrivare alla risoluzione di problemi, anche nuovi e mai incontrati prima (per i quali non vi è soluzione precedentemente programmata) sono già due caratteristiche che in certi casi potrebbero convincerci a dare a un software l'appellativo di intelligente. Ritroviamo infatti questi stessi elementi in molte delle definizioni moderne che, in questi ultimi anni, Stati, associazioni e organismi sovranazionali si sono affannati a redigere.

L'Unione Europea, grazie al gruppo di esperti che nel 2018 ha costituito lo High-Level Expert Group on Artificial Intelligence (HLEG-AI)[23], ha rilasciato forse una delle migliori definizioni contemporanee: "*I sistemi di intelligenza artificiale (IA) sono sistemi software (ed eventualmente anche hardware) progettati dall'uomo che, dato un obiettivo complesso, agiscono nella dimensione fisica o digitale percependo l'ambiente che li circonda attraverso l'acquisizione di dati, interpretando i dati strutturati o non strutturati raccolti, ragionando sulle conoscenze o elaborando le informazioni derivate da questi dati e decidendo la migliore o le*

[19] Armer Paul. *Attitude Toward Intelligent Machines* in Feigenbaum, A. *Computers And Thought*. New York: McGraw-Hill, 1963.
[20] https://www.britannica.com/technology/artificial-intelligence/Alan-Turing-and-the-beginning-of-AI.
[21] http://www.alanturing.net/turing_archive/archive/l/l32/L32-001.html.
[22] Somalvico, M. *Intelligenza Artificiale*. Milano: Scienza & Vita Nuova, 1997.
[23] https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence.

migliori azioni da intraprendere per raggiungere l'obiettivo prefissato. Sistemi di intelligenza artificiale possono utilizzare regole simboliche o imparare un modello numerico e possono anche adattare il loro comportamento analizzando come l'ambiente è stato influenzato dalle loro azioni precedenti."

Con questa definizione, che è praticamente la stessa adottata dal Governo italiano attraverso il gruppo di esperti del MiSE[24], non abbiamo quindi scorciatoie del tipo "sono intelligenti se sembrano intelligenti" né rimandi all'intelligenza umana. Del resto lo stesso HLEG-AI ha riconosciuto che la definizione di intelligenza - umana o meccanica - è ancora troppo vaga, nonostante sia stata ampiamente studiata da psicologi, biologi e neuroscienziati, preferendo quindi la nozione di razionalità, ossia la capacità di scegliere l'azione migliore per raggiungere un certo scopo.

Da segnalare che, più recentemente[25], il Parlamento europeo, riprendendo l'idea dell'intelligenza apparente, ha definito l'Intelligenza Artificiale come *un sistema basato su software o integrato in dispositivi hardware che mostra un comportamento intelligente, tra l'altro raccogliendo e trattando dati, analizzando e interpretando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere specifici obiettivi*.

Seguendo lo stesso filone del HLEG-AI, anche la definizione dell'OCSE, l'Organizzazione per la cooperazione e lo sviluppo economico, preferisce tralasciare i riferimenti all'intelligenza, prediligendo un approccio pragmatico e basato sul raggiungimento degli obiettivi[26]: "*Un sistema IA è un sistema basato su una macchina che può, per un dato insieme di obiettivi definiti dall'uomo, fare previsioni, raccomandazioni o decisioni che influenzano ambienti reali o virtuali. I sistemi di IA sono progettati per operare con diversi livelli di autonomia*". Da evidenziare, in questa definizione, un riferimento esplicito al concetto di autonomia che nelle altre definizioni o è sottinteso - perché si dà per scontato che una macchina intelligente debba saper operare autonomamente - oppure è ignorato - per evitare di complicare ulteriormente una situazione già di per sé intricata.

L'IA identificata come macchina risolutrice di problemi non è però un'idea recente. Raccontando la storia dell'intelligenza artificiale, nel 1995 Carlucci Aiello e Cialdea Mayer[27] spiegavano che "*lo studio dei giochi ha contribuito a centrare l'attenzione delle ricerche in intelligenza artificiale sulla soluzione automatica di problemi, riconoscendo come gran parte delle attività "intelligenti" possano essere viste come processi di soluzione di problemi. La ricerca in intelligenza artificiale si è caratterizzata infatti nel suo primo periodo come ricerca di algoritmi e di metodi generali e uniformi per la soluzione automatica di problemi.*"

Le difficoltà con cui in passato si sono confrontati ricercatori, professori e luminari, e che oggi stanno incontrando politici, giuristi e semplici funzionari nel definire qualcosa di sfuggente come l'intelligenza artificiale è una delle motivazioni - assieme alla proliferazione di

[24] https://www.mise.gov.it/images/stories/documenti/Proposte_per_una_Strategia_italiana_AI.pdf.
[25] Proposta di risoluzione recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate (2020/2012(INL) (A9-0186/2020), disponibile presso: https://www.europarl.europa.eu/doceo/document/A-9-2020-0186_IT.html#title1 .
[26] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.
[27] Carlucci Aiello, L. e Cialdea Mayer, M. *Invito all'intelligenza artificiale*. Franco Angeli, 1995.

formulazioni anche molto diverse fra loro - che hanno spinto gli autori di questo libro a non sceglierne nessuna in particolare, men che meno a crearne una propria. Dobbiamo inoltre tenere presente che l'IA è un concetto "aspirazionale" - negli anni Cinquanta con quel termine di certo non si descrivevano sistemi già esistenti, bensì teorie sperimentali molto in là dall'essere messe in pratica - difficile da definire anche perché costantemente *in fieri*. Una intelligenza artificiale, insomma, futura e futuribile, con il rischio concreto che il traguardo si sposti man mano che i ricercatori raggiungono nuovi risultati. Esiste infatti un principio formulato verso il 1970 noto come teorema di Tesler[28]: "*Intelligenza è tutto ciò che le macchine non hanno ancora fatto*", che il ricercatore Douglas Hofstadter ha parafrasato con "*IA è tutto quello che ancora non è stato fatto*", riassumendo il cosiddetto "effetto IA".

L'effetto IA è il costante declassamento da "intelligenza" a "mera computazione" che si fa quando un risultato, che prima si riteneva ottenibile solo attraverso l'intelligenza, viene infine raggiunto da una macchina. Come la vittoria del sistema di IBM Deep Blue sul campione di scacchi Garry Kasparov nel 1996 o quella di AlphaGo su Lee Sedol vent'anni dopo in un torneo di Go. Negli anni Sessanta qualcuno (ad esempio M. Taube[29]) affermava che le macchine non fossero intelligenti proprio perché non riuscivano a padroneggiare gli scacchi. Quando ciò è avvenuto si è preferito derubricare il traguardo come conseguito da un sistema basato su particolari algoritmi matematici, non ritenendo opportuno attribuire intelligenza ai software che hanno battuto i campioni umani. Come scriveva sempre Somalvico negli anni Ottanta, "*l'intelligenza artificiale [...] si occupa di problemi di ricerca aperti, di frontiera. Infatti una prestazione che oggi apparirebbe essere esclusiva dell'intelligenza umana, e quindi, secondo la definizione data, essere di pertinenza dell'IA, domani non creerebbe più questa impressione, lasciando l'osservatore perfettamente convinto della possibilità che la prestazione possa essere fornita da un sistema artificiale, e pertanto non rientrare più nell'intelligenza artificiale*".

Intervista a Luigia Carlucci Aiello, Professore Ordinario, cattedra di Intelligenza artificiale, Università Sapienza di Roma

Luigia Carlucci Aiello è stata Professore Ordinario dal 1981. Dal 1982 al DIS, ora DIAG, Sapienza Università di Roma; dal 1991 titolare della cattedra di Intelligenza artificiale. È stata coordinatrice del Dottorato di ricerca in Ingegneria informatica e Presidente del Consiglio di area in Ingegneria informatica. Dal 2006 al 2010 ha diretto il DIS, dal 2010 al 2013 è stata preside della neocostituita Facoltà di Ingegneria dell'Informazione, Informatica e Statistica. Fellow dell'AAAI (1995) e dell'ECCAI, ora EurAI (1999). Donald Walker Award dell'IJCAI (2009). Laurea Honoris Causa (2002) dalla Facoltà di Tecnologia dell'Università di Linköping – Svezia. ECCAI, ora EurAI Distinguished service Award (2014). Membro dell'Accademia Europea di Scienze e Arti (2015). Parte di questo testo è presente anche in "Robotica e

[28] http://www.nomodes.com/Larry_Tesler_Consulting/Adages_and_Coinages.html.
[29] Taube M. Letter to the editor. *Science*. 26 Agosto 1960 (tratto da Feigenbaum, A. *Computers And Thought*. New York: McGraw-Hill, 1963).

Intelligenza Artificiale: le sfide del secolo. La cultura italiana dal Novecento al nuovo millennio" a cura di Valerio Castelnuovo e Luigi Paparoni, Rizzoli 2020, pp 237-250.

Domanda 1. Se l'intelligenza artificiale in Italia oggi gode di solide basi e di un raro prestigio internazionale lo dobbiamo a lei e a quel gruppo di persone che vi ha lavorato quando molti la snobbavano. Potrebbe raccontarci brevemente come l'IA ha mosso i primi passi in Italia?

In Italia l'intelligenza artificiale ha mosso i primi passi già negli anni Cinquanta del secolo scorso. Tuttavia il nome intelligenza artificiale per molto tempo ha fatto fatica a trovare un suo spazio e una sua credibilità nella ricerca. Già nel 1958 Eduardo Caianiello costituì il Gruppo di Cibernetica presso l'Istituto di Fisica Teorica dell'Università di Napoli, il quale ha originato tanta della ricerca italiana sulle reti neurali. Negli stessi anni a Milano operava Silvio Ceccato - filosofo e linguista - e a Roma Vittorio Somenzi - filosofo della scienza, studioso degli automi e della relazione tra cibernetica e mente. Mentre al Politecnico di Torino iniziavano studi sulla comprensione della voce, al Politecnico di Milano iniziavano studi di robotica e, nel 1971, Marco Somalvico, di ritorno da un soggiorno presso il SAIL (Stanford Artificial Intelligence Laboratory) di Stanford (California - USA) vi fondò il Progetto di Intelligenza Artificiale e Robotica.

A partire dalla fine degli anni '50 a Pisa era iniziata la ricerca in informatica attorno alla CEP, Calcolatrice Elettronica Pisana, il primo computer costruito in Italia. Questa attirò l'attenzione dell'Olivetti (che creò a Pisa il famoso laboratorio di Barbaricina) e del mondo della ricerca internazionale, ma trovò inizialmente spazio solo in istituti CNR (il primo corso di laurea in Informatica in Italia sarebbe partito solo nel 1969, il primo corso di laurea in Ingegneria informatica addirittura nel 1990).

All'IEI (Istituto di elaborazione dell'informazione) del CNR di Pisa, negli anni '60 si faceva ricerca su elaborazione di immagini ed elaborazione simbolica, ma al reparto che si occupava di intelligenza artificiale era stato dato il nome di EINN, "Elaborazione dell'Informazione Non Numerica". Insieme a me ci lavoravano, tra gli altri, Ugo Montanari, Giorgio Levi, Franco Sirovich, Alberto Martelli, Antonio Albano e Mario Aiello.

Intorno agli istituti CNR pisani in quegli anni si muovevano anche Padre Roberto Busa, che, con la sponsorizzazione di Thomas Watson - fondatore dell'IBM - portò avanti con tenacia il progetto dell'Index Thomisticus, e Pietro Grossi che si dedicò a quella che oggi chiameremmo computer music.

Alla fine degli anni '70 il reparto EINN dell'IEI si è svuotato: praticamente tutti i suoi membri sono diventati professori ordinari e si sono sparsi per l'Italia. Io tra questi. Dopo la formazione pisana e due lunghi periodi in California al SAIL, dal 1982 ho fatto partire a Roma un gruppo di ricerca in intelligenza artificiale presso il Dipartimento ora denominato di Ingegneria Informatica Automatica e Gestionale della Sapienza. Negli anni il gruppo è cresciuto, fino ad

arrivare, col supporto di borse di studio industriali, borse di dottorato, e contratti sui primi progetti di ricerca europei, a una ventina di persone. Molte di queste ora sono professori in varie università italiane o lavorano in università e aziende in Europa e negli Stati Uniti. Ci siamo occupati di molti argomenti, la maggior parte riconducibili alla rappresentazione della conoscenza, al ragionamento automatico, alla costruzione di sistemi esperti, in particolare sistemi di supporto all'insegnamento, e più recentemente alla robotica cognitiva.

Anche far partire corsi di insegnamento in intelligenza artificiale nelle nostre università non è stato indolore. Il MIUR istituì la cattedra in Intelligenza artificiale nel 1990. Il primo corso fu tenuto da Somalvico al Politecnico di Milano nel '90, il secondo lo feci partire io nel '91 a Roma a Ingegneria alla Sapienza. Quell'anno lo seguirono 19 coraggiosi studenti, ma l'aula è presto diventata molto affollata. La tenacia e l'entusiasmo degli studenti hanno ripagato, ed ora fioriscono i corsi di insegnamento e i corsi di laurea in intelligenza artificiale e in robotica, ovvio prerequisito per la formazione delle nuove generazioni di ricercatori.

Domanda 2: Quali sono i centri di eccellenza nel nostro Paese?

Negli anni l'Italia ha partecipato in maniera attiva alla ricerca in intelligenza artificiale, collocandosi con forza nel panorama internazionale che agli inizi vedeva scambi e collaborazioni con gruppi nordamericani ed europei, poi si è estesa anche al resto del mondo. Risultati di rilievo ci hanno portato all'attenzione internazionale nel campo della cibernetica, dell'automatica, dello sviluppo di robot per applicazioni industriali, in ambienti ostili o pericolosi, delle tecniche di rappresentazione della conoscenza e del ragionamento automatico, della comprensione del linguaggio parlato e scritto e della visione. La ricerca è stata portata avanti in molte università e politecnici italiani (Torino, Milano, Genova, Padova, Bologna, Pisa, Roma, Napoli, Bari, Palermo), in istituti CNR e in istituti di ricerca quali lo CSELT a Torino, dove sono state sviluppate tecnologie per la comprensione del parlato, o l'IRST dell'ITC, ora Fondazione Kessler, a Trento, dove Luigi Stringa – che aveva già sviluppato presso ELSAG la macchina EMMA per il riconoscimento dei codici di avviamento postale per le poste italiane – a metà degli anni '80 propose di fare ricerca in intelligenza artificiale e costruire robot mobili autonomi. A Genova, dove la ricerca in intelligenza artificiale e robotica era già attiva da molti anni sia all'università che al CNR, è stato istituito nel 2003 l'IIT, Istituto Italiano di Tecnologia, che si è successivamente esteso in molte sedi in Italia e all'estero e dalla cui ricerca sono emersi interessanti prototipi e risultati. A Pisa, dalla ricerca svolta in università e al CNR, si è sviluppata la ricerca in robotica, in particolare in biorobotica ha trovato molto spazio alla Scuola Superiore Sant'Anna, e alla sua estensione nel Polo Sant'Anna Valdera. Potrei continuare l'elenco: ci sono altri centri in Italia dove si fa ricerca ad alto livello, ma non voglio dilungarmi troppo, in particolare non mi dilungo a parlare ulteriormente della ricca e variegata realtà romana divisa in vari dipartimenti delle tre università Statali, in vari istituti CNR e non solo.

Nel 1988, insieme a un gruppo di colleghi operanti in università, centri di ricerca e aziende ho fondato – e presieduto per quattro anni l'AIxIA, Associazione Italiana per l'Intelligenza

Artificiale. Molto impegnata nel favorire la formazione, la diffusione della ricerca e la collaborazione tra i vari gruppi, l'associazione ha da poco compiuto trent'anni e continua la sua attività con la presidenza del Dottor Piero Poccianti. L'AIxIA rappresenta l'Italia dentro l'EurAI, l'associazione europea per l'intelligenza artificiale, e vi svolge un importante ruolo. Ho appena contato i Fellow italiani nell'EurAI: 20 operanti in istituzioni Italiane, 5 operanti all'estero. Niente male come riconoscimento della qualità del lavoro svolto dagli italiani, se si considera la dimensione paneuropea dell'EurAI e che il numero totale degli EurAI Fellow è attualmente circa 170.

Non trascurabile neanche il prestigio di cui godono gli italiani all'interno dell'AAAI, Association for the Advancement of Artificial Intelligence, associazione di dimensione mondiale, nata nel 1980 e fortemente incentrata nel nord America. In essa sono presenti 4 Fellows italiani operanti in Italia e 3 operanti all'estero. Qui le liste sono più brevi e si possono elencarne i nomi. La prima lista, in ordine di tempo: Luigia Carlucci Aiello (Roma), Oliviero Stock (Trento), Giuseppe De Giacomo (Roma) e Maurizio Lenzerini (Roma); la seconda: Maria Gini (USA), Francesca Rossi (USA) e Marco Dorigo (Belgio). Con una punta di orgoglio lasciatemi anche aggiungere che nel primo gruppo tre su quattro sono del mio dipartimento e nel secondo la Francesca Rossi è stata recentemente eletta vicepresidente, e quindi diventerà presidente, dell'AAAI stessa.

Domanda 3: Sembra che il deep learning abbia regalato all'intelligenza artificiale una nuova primavera. Cosa dobbiamo fare affinché non torni un altro inverno?

Avendo lavorato per circa 50 anni in intelligenza artificiale posso dire di aver visto tante stagioni susseguirsi. Tante diatribe, tante lotte tra gruppi che tentavano di imporre un metodo, una tecnica, o solamente un punto di vista su un altro. I più vecchi del campo ricordano le eterne lotte su logica o altri formalismi per la rappresentazione della conoscenza, sull'approccio dichiarativo o procedurale alla soluzione automatica dei problemi, sull'approccio simbolico o subsimbolico. Che dire della battaglia attorno ai linguaggi general purpose o linguaggi per l'intelligenza artificiale? E il ProLog, in particolare dopo che fu scelto dai giapponesi come linguaggio per le macchine di quinta generazione? L'acceso dibattito intorno al ProLog portò Alan Robinson - il creatore dell'algoritmo di unificazione che ne è il cardine - a dire "Io ho proposto l'unificazione come algoritmo, non come way of life".

Ovviamente ogni volta che una tecnica si è dimostrata promettente semplicemente perché con essa è stato risolto un problema comprensibile ai più, considerato molto difficile e non risolto con tecnologie informatiche in precedenza, si è parlato di primavera. O meglio si è esultato, salvo poi parlare di inverno quando l'entusiasmo scemava di fronte a qualche grosso scoglio. Ne ho visti vari di inverni di solito legati al tentativo di mettere (a caro prezzo) sul mercato, spacciandoli per consolidate soluzioni, strumenti che avevano solo la robustezza di un prototipo di laboratorio. Questo, secondo me, va accuratamente evitato: non c'è un'intelligenza artificiale buona e una cattiva, ci sono solo sistemi che funzionano bene o funzionano male, nel secondo caso non sono pronti a essere messi in circolazione.

Quindi ben venga l'entusiasmo per le reti neurali multistrato, e anche le tecniche di apprendimento statistico sono le benvenute. Recentemente mi tornava in mente una proposta che avevo fatto ad una grossa azienda per un progetto di loro forte interesse interno negli anni '80. Il mio interlocutore aziendale la bocciò, anche se la considerava una idea brillante, perché avrebbe richiesto una pesante pre-elaborazione manuale dei dati a loro disposizione che secondo lui la rendeva inapplicabile. Beh, oggi la pre-elaborazione si farebbe in un fiat.

Per usare uno slogan, io appartengo alla GOFAI (Good Old Fashioned Artificial Intelligence), ma ho un grande rispetto per la ricerca sul machine learning e sui risultati che si raggiungono con tecniche di deep learning. Sottoscrivo che in questi ultimi tempi i risultati più importanti sono arrivati in settori in cui la GOFAI poco aveva potuto, e non è dall'utilizzo del deep learning per sé che mi aspetto il prossimo inverno.

Certo potrebbe tornare un altro inverno dell'intelligenza artificiale, se le aspettative fossero troppo alte a fronte di ricerca e/o investimenti inadeguati. Dobbiamo sempre pensare che di ricerca stiamo parlando, che i problemi difficili esistono e non si risolvono solo adottando parole di moda e altisonanti come lampade magiche in mano a geni che le usano pronunciando "apriti sesamo". Quindi, forse, il rischio più grosso di questa calda primavera dell'intelligenza artificiale sta proprio nei troppi geni con in mano la lampada.

Stiamo attraversando un periodo in cui l'economia mondiale, la salute, l'ambiente sono in crisi. Le tecnologie digitali che oggi vengono viste sotto l'unico cappello dell'intelligenza artificiale e identificate con un algoritmo che fa apprendimento da grandi quantità di dati possono giocare un grande ruolo in tutti questi settori e salvare l'umanità dalle calamità incombenti. Basta non combattere guerre parrocchiali o pensare che l'era della ricerca è conclusa e adesso siamo allo sviluppo e alle applicazioni di routine.

6 I componenti dell'IA

In termini di architetture hardware e software, le soluzioni basate su IA comprendono solitamente i seguenti componenti.

- I **dati** di partenza (spesso presenti in grandi quantità, e quindi rientranti nella categoria dei big data), che vengono analizzati tramite algoritmi statistici, per individuare schemi, regolarità, anomalie, ecc.
- L'**edge**, dove i dati possono essere analizzati e in parte elaborati prima dell'invio ai servizi di calcolo presenti nel cloud.
- Il **cloud**, dove vengono svolte le elaborazioni più complesse.
- Le **applicazioni di IA**, che possono essere fornite agli utilizzatori attraverso un sito web. Quasi sempre, oltre all'interfaccia web vengono fornite delle API grazie alle quali gli sviluppatori software possono integrare gli algoritmi di IA all'interno di loro applicazioni.
- Le **applicazioni desktop** per PC e per **dispositivi mobili** usate per collegarsi tramite API a un servizio di IA.

La crescita della capacità elaborativa dei microprocessori e delle piattaforme elettroniche a basso costo ha permesso di portare l'intelligenza artificiale dal cloud verso la periferia, verso il piccolo e il decentrato, permettendo applicazioni ibride, locali o addirittura personali. Come esempio si può fare riferimento agli assistenti vocali per smartphone, che ascoltano la voce del proprietario, la pre-elaborano in locale e mandano il risultato di tale pre-elaborazione ai server nel cloud, che esaminano la domanda posta e forniscono una risposta appropriata.

Un altro interessante esempio, relativo anche alle capacità di elaborazione dei sensori periferici, sono i moderni sistemi di guida autonoma, con livelli di autonomia dal livello 2 al livello 5 (vedere paragrafo 15.2). Anche in ambito industriale si usa intelligenza e aggregazione locale (con componenti edge e IoT): le componenti periferiche raccolgono i dati ed effettuano una prima pulizia dei dati prima che essi siano inviati ai sistemi centrali; questo riduce il carico elaborativo centrale e migliora le prestazioni complessive. Per questo, recentemente sono stati lanciati sul mercato dispositivi edge ad altissime prestazioni, veri e propri server locali dotati di grandi capacità di elaborazione e filtraggio.

Oltre ad una disamina più "fisica" delle componenti proprie di una soluzione di IA, come fatto sopra, si può pensare di condurre un'analisi più legata ai flussi logici. Da questo punto di vista, i sistemi di intelligenza artificiale integrati sono composti da quattro componenti essenziali, dedicati ad altrettante funzionalità logiche: percezione, apprendimento, decisione e azione (nel paragrafo 15.2 queste funzionalità sono presentate nell'ambito dei veicoli a guida autonoma).

7 Algoritmi e strumenti teorici

Nei prossimi paragrafi sono approfonditi alcuni concetti e nozioni utilizzati nell'ambito dell'IA. Vista la natura divulgativa di questa pubblicazione, per ciascuno degli argomenti trattati è data una descrizione il più possibile precisa - ma senza utilizzare formalismi di tipo matematico - cercando di privilegiare la chiarezza rispetto alla completezza o alla precisione formale.

7.1 Le 5 tribù dell'IA

Iniziamo questo capitolo parlando di cinque approcci - o scuole di pensiero - che si sono rivelati molto efficaci per la risoluzione di problemi di varia natura e per la progettazione di algoritmi di IA. Come indicato in un popolare libro di Pedro Domingos[30], i ricercatori che si occupano di studiare questi algoritmi hanno formato nel corso del tempo cinque vere e proprie "tribù", ciascuna concentrata sullo studio delle caratteristiche di una particolare classe di algoritmi, e dei problemi da essi risolti:

- I **simbolisti** traggono ispirazione dalla filosofia e dal ragionamento logico basato sull'astrazione dei simboli. Chi fa parte di questo gruppo si occupa, ad esempio, di *rappresentazione della conoscenza*, tramite *ontologie* e *database di conoscenza*, e di algoritmi di *ragionamento deduttivo* e *induttivo*. I metodi per la rappresentazione della conoscenza consentono di rappresentare **entità** (fatti, persone, ecc.) e **relazioni** tra esse; la rappresentazione deve essere efficiente, sia per l'interrogazione delle relazioni esistenti sia per consentire l'aggiunta di nuovi fatti, comprendenti eventualmente nuove entità e relazioni. L'esempio più famoso di questo tipo di rappresentazione della conoscenza è il *grafo di conoscenza* (*knowledge graph*) di Google, di cui viene mostrato un frammento in ogni pagina di risultati di una ricerca. Gli algoritmi di ragionamento, implementati solitamente sotto forma di *sistemi di regole* o di *alberi di decisione*, possono essere visti come una sequenza di **passi di deduzione**, ciascuno dei quali trae una conclusione a partire da certe ipotesi, costituite da fatti e relazioni note. Oppure, viceversa, possono essere visti come algoritmi che cercano di capire se a partire da certe ipotesi è possibile dedurre un fatto o una relazione desiderati, tramite un'opportuna sequenza di passi di deduzione; si parla in questo caso di **deduzione inversa**. L'*IA simbolica* è stata il paradigma dominante tra gli anni 1950 e 1980, e ha prodotto molti software interessanti, tra cui i cosiddetti *sistemi esperti*[31].

[30] Pedro Domingos. *L'Algoritmo Definitivo. La macchina che impara da sola e il futuro del nostro mondo*. Bollati Boringhieri, 2016.

[31] Nicoletta Boldrini. Sistemi esperti: cosa sono, la loro classificazione, come funzionano e a cosa servono. *AI4Business.* 23 settembre 2019. Disponibile su: https://www.ai4business.it/intelligenza-artificiale/sistemi-esperti-cosa-sono/

- I **connettivisti**, o **connessionisti**, si basano sulle neuroscienze e traggono ispirazione da come è fatto e come funziona il cervello umano a livello di reti di neuroni (e quindi, in particolare, su come i neuroni sono connessi e comunicano tra loro). I membri di questa tribù studiano le *reti neurali artificiali* nelle loro innumerevoli varianti, compreso l'ormai famoso *deep learning*. Come vedremo nel paragrafo dedicato, i modelli di reti neurali possono essere più o meno aderenti al funzionamento delle reti di neuroni biologiche, ma in ogni caso sono dotate della capacità di apprendere a partire da esempi, modificando le connessioni tra un neurone e l'altro.
- Gli **evoluzionisti** applicano i principi della biologia evolutiva, in particolare le operazioni svolte dal DNA durante la riproduzione degli individui, nonché la nozione di *fitness*, una sorta di misura di quanto l'individuo si è adattato alle condizioni dell'ambiente in cui vive. Gli algoritmi di questo tipo fanno evolvere delle popolazioni di possibili soluzioni per problemi, che solitamente richiedono di massimizzare o minimizzare certe quantità, soddisfacendo contemporaneamente alcuni vincoli. Usando opportune operazioni di incrocio (*crossover*) e di mutazione casuale, gli individui della popolazione producono la successiva generazione di individui, che si spera siano più *adatti* a risolvere il problema. Oltre che per risolvere diversi problemi di ottimizzazione, gli algoritmi evolutivi possono essere usati per far evolvere popolazioni di *reti neurali artificiali*, oppure di *programmi per computer*.
- I **bayesiani** adottano un approccio dettato dalla statistica e dal calcolo delle probabilità per implementare una forma di inferenza (ragionamento) probabilistica. Come vedremo, gli algoritmi di questa categoria assumono la forma di *reti causali*, o *bayesiane*, in cui vengono valutate le probabilità (condizionate) che si verifichino certi eventi a partire da certe ipotesi. Una volta stabilite le relazioni probabilistiche tra cause ed effetti, l'utilizzo del teorema di Bayes consente di calcolare la probabilità che una causa abbia scatenato un certo evento dato. L'applicazione più famosa di questo tipo di algoritmi sono i *filtri antispam* per le e-mail.
- Gli **analogisti**, i cui algoritmi stabiliscono associazioni tra dati, calcolate usando le cosiddette *support vector machines* (SVM). L'esempio più famoso di algoritmi in questa categoria sono i cosiddetti *recommender systems*, che in base alle preferenze mostrate dagli utenti suggeriscono un prodotto da acquistare, un film da vedere o una canzone da ascoltare. Ciò avviene perché, per l'appunto, vengono stabilite delle associazioni - e calcolate delle somiglianze - tra i dati da noi generati (prodotti acquistati, film visti, ecc.) e le scelte fatte precedentemente da chi ha mostrato di avere gusti simili ai nostri.

Nei prossimi paragrafi verranno descritte - in maniera pur sempre introduttiva, senza poter scendere nei dettagli - alcune di queste tipologie di algoritmi. Qui vogliamo invece fare due osservazioni. La **prima osservazione** è che questi cinque approcci, in realtà, non esauriscono tutte le possibilità dell'intelligenza artificiale, che si occupa anche di algoritmi per la visione, per il riconoscimento del linguaggio naturale, per la gestione cognitiva di corpi robotici, e altro.

La **seconda osservazione** è più profonda e molto più interessante e risponde alle seguenti domande: "Dato un problema, qual è il tipo di algoritmo più adatto?", e "Quale tipo di algoritmo è più performante?" In termini astratti *ciascuna* delle cinque categorie menzionate sopra è in grado di risolvere *qualsiasi* tipo di problema ed è candidata a contenere un **algoritmo universale**, cioè un algoritmo che sia in grado di risolvere qualsiasi problema di apprendimento si possa concepire. Ma allora, perché continuare a considerare tutte e cinque

le categorie di algoritmi menzionate sopra, anziché concentrarsi su una sola? La risposta è che, mentre questo può avvenire in teoria, in pratica la quantità di dati necessari per addestrare un algoritmo per risolvere un problema dato può dipendere fortemente sia dal tipo (cioè della categoria a cui appartiene) di algoritmo, sia dal tipo di problema. Per ogni tipo di problema esistono algoritmi più adatti o più efficienti e algoritmi meno adatti o meno efficienti per risolverlo. Il caso limite è dato dal fatto che, per certi problemi, la quantità di esempi necessari per addestrare gli algoritmi di certe categorie a risolvere tali problemi può tendere all'infinito.

Tuttavia, queste considerazioni non hanno scoraggiato i ricercatori, che stanno continuando a cercare l'algoritmo universale, chiamato a volte anche l'**algoritmo definitivo**. Un'ipotesi molto plausibile è che l'algoritmo universale, se esiste, potrebbe essere una combinazione di tutti gli algoritmi appartenenti alle cinque categorie viste sopra. Ma quali caratteristiche avrebbe tale algoritmo, se venisse scoperto? Oltre a essere in grado di risolvere *qualsiasi* problema usando una quantità ragionevole di esempi per l'apprendimento, l'algoritmo definitivo sarebbe in grado di estrarre *tutte* le informazioni possibili dai dati, e fare *qualsiasi cosa* con tali informazioni. Sarebbe così in grado, ad esempio, di automatizzare le scoperte, progettando ed eseguendo esperimenti in completa autonomia, cioè in maniera completamente *non supervisionata*. Implementato su un hardware opportuno, costituirebbe una macchina che impara da sola, dalle proprie esperienze. Analizzando il proprio programma (cioè l'implementazione dell'algoritmo universale), tale macchina sarebbe in grado di progettarne versioni più efficienti, migliorando quindi se stessa. Il tutto avverrebbe a velocità sempre maggiori, dovute sia (banalmente) all'aumento della velocità dell'hardware, sia al miglioramento delle prestazioni nelle successive versioni dell'algoritmo, che richiederebbero meno dati e meno istruzioni di elaborazione per ogni dato rispetto alle versioni precedenti. L'algoritmo universale diventerebbe appunto l'*algoritmo definitivo*, perché sarebbe l'ultima invenzione necessaria da parte degli esseri umani: dal momento in cui l'algoritmo dovesse entrare in funzione, prenderebbe in mano lui il controllo delle invenzioni e delle scoperte nell'ambito dell'IA (e non solo). Ciò ha naturalmente a che fare con la teoria dell'intelligenza artificiale generale (AGI), detta a volte anche "intelligenza artificiale forte", in grado di fare tutto quello che può fare il cervello umano, e anche più (o anche solo più velocemente).

Come nota finale, è opportuno precisare che, a volte, nell'ambito del machine learning, al posto di *algoritmo universale* si usa la locuzione **modello universale**. L'idea è la medesima, ristretta al fatto che nel machine learning si parla di *addestramento di modelli*, anziché della (potenzialmente più generica) *formulazione di algoritmi*.

7.2 Modalità di apprendimento

Gli algoritmi di machine learning (in italiano: apprendimento automatico), oltre a quanto visto nel paragrafo precedente, possono anche essere classificati sulla base delle diverse modalità di apprendimento utilizzate, come accennato in precedenza; ciascuna si differenzia per le caratteristiche dell'insieme di dati (*dataset*) usato per l'addestramento. Nel seguito sono proposte tre famiglie.

- Si parla di **apprendimento supervisionato** quando i dati utilizzati per l'apprendimento sono costituiti da input (valori delle cosiddette *features*) e da output attesi. Per esempio, un sistema di riconoscimento delle immagini viene addestrato fornendogli immagini con già assegnata un'etichetta che rappresenta il contenuto. Questi dati agiscono quindi come un insegnante che supervisiona uno studente durante l'apprendimento. Tali algoritmi, terminato l'addestramento, sono in grado di risolvere problemi in modo autonomo basandosi sull'esperienza formatasi sugli esempi forniti precedentemente e indicati come "appropriati". Una tipologia di apprendimento supervisionato è la **classificazione,** dove il sistema deve produrre un modello in grado di assegnare a un input una o più classi tra quelle disponibili. Altro esempio di apprendimento supervisionato è la **regressione**, che permette di analizzare le relazioni esistenti tra le variabili indipendenti e le variabili dipendenti di un problema, ad esempio cercando la curva che meglio approssima le osservazioni empiriche fatte. Uno degli ambiti storici di applicazione dell'apprendimento supervisionato alla cybersecurity è rappresentato dalla rilevazione delle mail di spam, dove vennero originariamente utilizzati dei semplici classificatori bayesiani. Attualmente gli algoritmi supervisionati sono utilizzati come moduli all'interno di sistemi più complessi, ad esempio per la classificazione del traffico di rete o del contenuto delle pagine web. Un'altra applicazione è quella adottata per stabilire (classificare) se un'app Android contiene malware[32], anche se tale compito è estremamente difficile poiché ci sono notevoli difficoltà - sia di tipo teorico che di natura pratica - nell'imparare che cos'è un malware.
- Si parla di **apprendimento non supervisionato** quando i dati forniti contengono solo gli input, senza nessun output atteso. Tali algoritmi apprendono in autonomia la struttura (nascosta) dei dati: agiscono quindi come studenti che apprendono senza la supervisione di un insegnante. Un esempio di apprendimento non supervisionato è il **clustering**, in cui un insieme di dati viene raggruppato in cluster (gruppi) in modo che i dati nello stesso gruppo abbiano caratteristiche simili e dati in gruppi distinti abbiano caratteristiche molto diverse (vedere paragrafo 3.3).
- Nell'**apprendimento con rinforzo** (reinforcement learning) non vengono forniti dati di training; in questo caso il sistema interagisce con l'ambiente nel quale è immerso. Il sistema sceglie una tra tante azioni possibili e la risposta dell'ambiente a tale azione comprende un segnale di ricompensa o di penalità; il sistema quindi apprende cercando di massimizzare la ricompensa ottenuta. Un esempio classico di applicazione dell'apprendimento con rinforzo è il gioco degli scacchi: quando tocca al computer muovere, per ogni mossa possibile viene valutata la configurazione risultante sulla scacchiera (in realtà, questa valutazione viene fatta considerando un certo numero di mosse e contromosse da parte dell'avversario); viene infine scelta una delle mosse che massimizzano il vantaggio risultante, cioè le possibilità di vincere la partita. Il reinforcement learning è diventato famoso per imparare a giocare a go (grazie al progetto AlphaGo[33]) e, recentemente, ad alcuni videogiochi della console Atari 2600[34] - in alcuni casi meglio degli esseri umani.

[32] K. Allix et al., Empirical assessment of machine learning-based malware detectors for Android. *Empirical Software Engineering* 21, 183–211 (2016). https://doi.org/10.1007/s10664-014-9352-6.
[33] https://deepmind.com/research/case-studies/alphago-the-story-so-far.
[34] V. Mnih et al., Playing Atari with Deep Reinforcement Learning (2013). arXiv:1312.5602v1.

I sistemi basati sull'apprendimento sono caratterizzati da altre due proprietà significative. La prima è la capacità di migliorare le proprie prestazioni nel tempo, in considerazione del fatto che l'algoritmo funziona tanto meglio quanto più elevato è il volume di dati disponibile. La seconda è rappresentata dalla capacità dell'algoritmo di "adattarsi" ai dati massimizzando le proprie prestazioni rispetto a un particolare contesto di riferimento. Due esempi classici, in tal senso, sono i filtri anti-spam e i sistemi di riconoscimento vocale: in entrambi i casi, l'algoritmo viene prima modellato su un insieme di dati dimensione molto elevato e relativo a molti utenti e poi specializzato sui dati dello specifico utente (la sua voce, nel caso dei sistemi di riconoscimento vocale, o le sue risposte al filtro anti-spam nell'altro caso).

Il fatto che le prestazioni del sistema migliorino con un volume elevato di dati (i cosiddetti big data) è reso possibile - dal punto di vista tecnologico - dai progressi nella capacità di immagazzinare i dati (storage più capiente) e dall'ampia disponibilità di applicativi e dispositivi attraverso cui i dati possono essere raccolti ed elaborati. Le basi di dati utilizzate per allenare le IA traggono inoltre beneficio dalla presenza di modelli di business in cui i dati vengono ceduti gratuitamente dagli utenti in cambio di servizi gratuiti come motori di ricerca, email, traduzione di testi e dettatura di messaggi.

7.3 Classificazione e clustering

Approfondiamo ora una tecnica statistica, il *clustering*, che può essere utilizzata per produrre sia gli omonimi algoritmi di apprendimento *non supervisionato* - menzionati nel paragrafo precedente - sia algoritmi di *classificazione*. Il fatto che questi ultimi siano stati classificati nel paragrafo precedente come algoritmi di apprendimento *supervisionato* non deve confondere il lettore: in IA succede spesso che le buone idee vengano riutilizzate, anche in ambiti diversi. In questo caso i raggruppamenti di dati prodotti dal clustering possono essere utilizzati come "indicazioni" o "etichettature" per i valori desiderati in output durante l'addestramento (supervisionato) di un algoritmo di classificazione.

In statistica, il *clustering* è un insieme di tecniche per l'individuazione di raggruppamenti omogenei degli elementi che costituiscono una popolazione di dati.

A partire dalla rilevazione degli attributi, quantitativi o qualitativi, propri degli elementi in osservazione, le tecniche di clustering applicano metriche in grado di calcolare la *distanza* - spesso definita come **similarità** - tra gli elementi che, di conseguenza, tendono a costituire dei gruppi. Il processo può quindi portare a una *semplificazione* nella percezione delle caratteristiche della popolazione osservata, in modo da determinare la classificazione dei suoi elementi e fissare le caratteristiche di ogni cluster, abilitando così la possibilità di riconoscere ed incasellare nuovi elementi nel modello.

Gli algoritmi di clustering si dividono in due grandi categorie.

- **Algoritmi agglomerativi**, detti anche "bottom-up": considerano ogni elemento in osservazione come un potenziale cluster e applicano tecniche iterative di misurazione per raggiungere le condizioni predefinite come, per esempio, il numero di cluster desiderati.

- **Algoritmi divisivi**, detti anche algoritmi gerarchici o top-down: si tratta di algoritmi che inizialmente mettono tutti gli elementi in un unico cluster; dopodiché dividono gli elementi raggruppandoli in sotto-cluster (normalmente creando un numero fisso di cluster alla volta) in base a diversi possibili criteri. Solitamente, i cluster da dividere vengono scelti in base a funzioni che misurano la *compattezza* del cluster e la *densità* o la *sparsità* degli elementi assegnati a un cluster; in ogni caso, l'idea è che gli elementi appartenenti a un cluster dovrebbero essere più simili tra loro rispetto a elementi appartenenti a cluster diversi. Le operazioni di divisione vengono reiterate fino a raggiungere il numero desiderato di cluster.

Non è però detto che la popolazione, per le sue caratteristiche intrinseche e gli attributi rilevati, possa essere ricondotta a raggruppamenti, o che i raggruppamenti individuati rispecchino la distribuzione naturale dei dati, o che questi si presentino con caratteristiche sufficientemente nette da poter essere effettivamente utilizzati.

Di conseguenza, risulta spesso necessario tentare più volte e applicare diverse tecniche: si possono cambiare le metriche, si può ragionare per algoritmi agglomerativi oppure divisivi, si possono applicare tecniche esclusive per cui un elemento deve appartenere a un solo cluster, o non esclusive per cui un elemento può appartenere a più cluster.

Un esempio di applicazione del clustering è il *role engineering*. Esso permette di stabilire un modello di assegnazione delle abilitazioni sui sistemi IT aziendali basato sull'identificazione di cluster omogenei di ruoli attribuibili al personale in base alla collocazione in organigramma o alla mansione ricoperta. L'algoritmo di clustering, in questi casi, viene alimentato dalle informazioni sulla struttura organizzativa dell'organizzazione e dai permessi attribuiti alle persone e configurati sui sistemi IT, applicando un approccio cosiddetto *ibrido*, che combina analisi agglomerative e divisive.

Un ulteriore esempio di applicazione del clustering è la generazione delle firme per la rilevazione dei malware. I malware sono raggruppati sulla base delle loro caratteristiche comportamentali (p.e. caratteristiche del traffico generato); tali comportamenti possono essere raggruppati in cluster per poter identificare il malware che li genera. Considerato che il processo può essere particolarmente oneroso per via della numerosità dei campioni (possono essere raccolti diversi milioni di campioni di malware in una sola giornata), vengono alternate fasi in cui i nuovi campioni, via via che vengono reperiti, vengono aggiunti a cluster esistenti, e fasi in cui l'intero dataset di malware a disposizione viene utilizzato per produrre un nuovo clustering, in quanto è necessario tener conto della mutevolezza, nel tempo, della distribuzione dei dati.

https://www.youtube.com/watch?v=mJeNghZXtMo

“What is Artificial Intelligence (or Machine Learning)?”

Fornisce una descrizione di massima dei termini machine learning e IA.

(durata: 6 minuti e 14 secondi)

https://www.youtube.com/watch?v=ad79nYk2keg&list=PLEiEAq2VkUULyr_ftxpHB6DumOq1Zz2hq

“Artificial Intelligence In 5 Minutes”

Questo video ti introduce sull’Intelligenza Artificiale con un approccio semplice e divertente.

(durata: 5 minuti e 27 secondi)

https://www.youtube.com/watch?v=iEGE0nxPnnA

“Machine Learning: Le Basi (Intelligenza Artificiale)”

Come funziona il machine learning? Partiamo dalle basi.

(durata: 15 minuti e 18 secondi)

https://www.youtube.com/watch?app=desktop&v=U78wETHkBKU&list=PLa-sizbCyh93evwIevvnjWFEH94N5giIG.

“Machine learning in Python”

Corso di machine learning in python e scikit-learn.

(durata: 5 minuti e 23 secondi)

7.4 Approcci allo sviluppo dell'IA

Approfondiamo infine alcune tipologie di algoritmi di IA, che hanno riscosso un notevole successo nell'applicazione a problemi reali. Come vedremo, questi algoritmi sono nati seguendo approcci abbastanza diversi tra loro e applicando tecniche matematiche altrettanto diverse. Ritroveremo molti degli algoritmi menzionati nel paragrafo 7.1; tuttavia, mentre in tale paragrafo ci eravamo concentrati sulle "scuole di pensiero" seguite dai ricercatori di IA, in questo paragrafo il punto di vista sarà quello dei modelli computazionali e delle tecniche matematiche usate per progettare gli algoritmi di apprendimento.

7.4.1 L'approccio connessionista

L'approccio connessionista all'intelligenza artificiale propone algoritmi e modelli computazionali che si ispirano al modo in cui le reti neurali biologiche del cervello elaborano le informazioni. Il cervello umano è costituito da neuroni, cellule nervose interconnesse che si occupano dell'elaborazione e della trasmissione di segnali chimici ed elettrici. La prima idea, messa a punto negli anni '40 del secolo scorso da Warren McCulloch e Walter Pitts, è di rappresentare un neurone biologico come una sorta di porta logica con diversi input e un output, ciascuno dei quali può assumere valore 0 oppure 1; inoltre, ciascun input può essere *eccitatorio* o *inibitorio*. Il neurone calcola un *valore aggregato* dei valori di input, secondo una funzione prestabilita; se il risultato supera una soglia prefissata viene prodotto in output un 1, in caso contrario viene prodotto uno 0.

Il primo **algoritmo di apprendim ento** basato su questa rappresentazione di neurone è stato messo a punto nel decennio successivo da Frank Rosenblatt, che ebbe l'idea di associare agli input del neurone dei *pesi* variabili positivi (azioni eccitatorie) o negativi (azioni inibitorie), dando vita al *percettrone*, detto anche *neurone binario a soglia*.

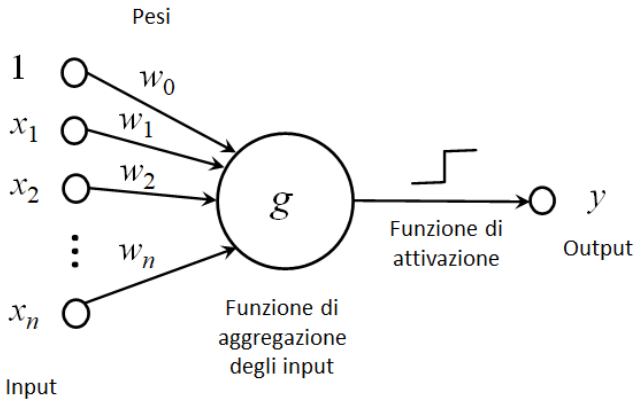


Figura 1 - Il percettrone[35]

[35] Fonte: figura autoprodotta.

La prima implementazione (hardware) di percettrone, realizzata da Frank Rosenblatt, fu chiamata *Mark I perceptron* (si veda la Figura 2).

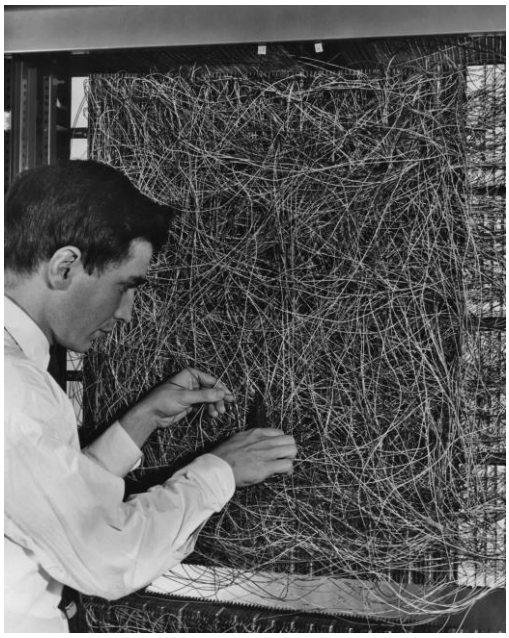


Figura 2 - Frank Rosenblatt al lavoro sul Mark I perceptron[36]

L'algoritmo proposto permette l'apprendimento variando i coefficienti di peso e la soglia di attivazione in modo da minimizzare la differenza tra l'output ottenuto e quello desiderato. I problemi per cui l'algoritmo del percettrone può essere utilizzato sono quelli supervisionati di classificazione binaria. Una grossa limitazione, evidenziata da Marvin Minsky e Seymour Papert[37], è che l'algoritmo di apprendimento converge solo nel caso in cui le classi a cui appartengono i dati in input sono *linearmente separabili*. Purtroppo esistono problemi - anche molto semplici - che non lo sono e che quindi non sono affrontabili con un solo percettrone.

Questa limitazione può essere superata collegando fra loro diversi strati di neuroni binari a soglia, ottenendo le cosiddette **reti neurali artificiali** (*artificial neural networks*) multistrato. In tali reti i neuroni del primo strato acquisiscono i valori di input, li elaborano e passano i risultati ai neuroni dello strato successivo, i quali elaborano i valori ricevuti e passano i risultati allo strato successivo, e così via; l'ultimo strato è composto da uno o più neuroni di output. Il fatto che negli anni '60 non si sapesse come addestrare questo tipo di reti neurali ha fatto subire alla ricerca e alla sperimentazione sulle reti neurali una battuta d'arresto fino all'inizio degli anni '80, quando si è capito come applicare la tecnica matematica del *gradiente discendente* alla modifica dei pesi della rete, a partire dai pesi dei neuroni di output e andando "all'indietro", verso i pesi dei neuroni di input, in modo da diminuire la differenza tra l'output ottenuto e quello desiderato. Nasce così l'**algoritmo di retropropagazione** dell'errore.

[36] Fonte: https://www.nzz.ch/digital/ehre-fuer-die-deep-learning-mafia-ld.1472761 (Bild Getty Images).
[37] Marvin Minsky, Seymour Papert. *Perceptrons*. MIT Press, 1969.

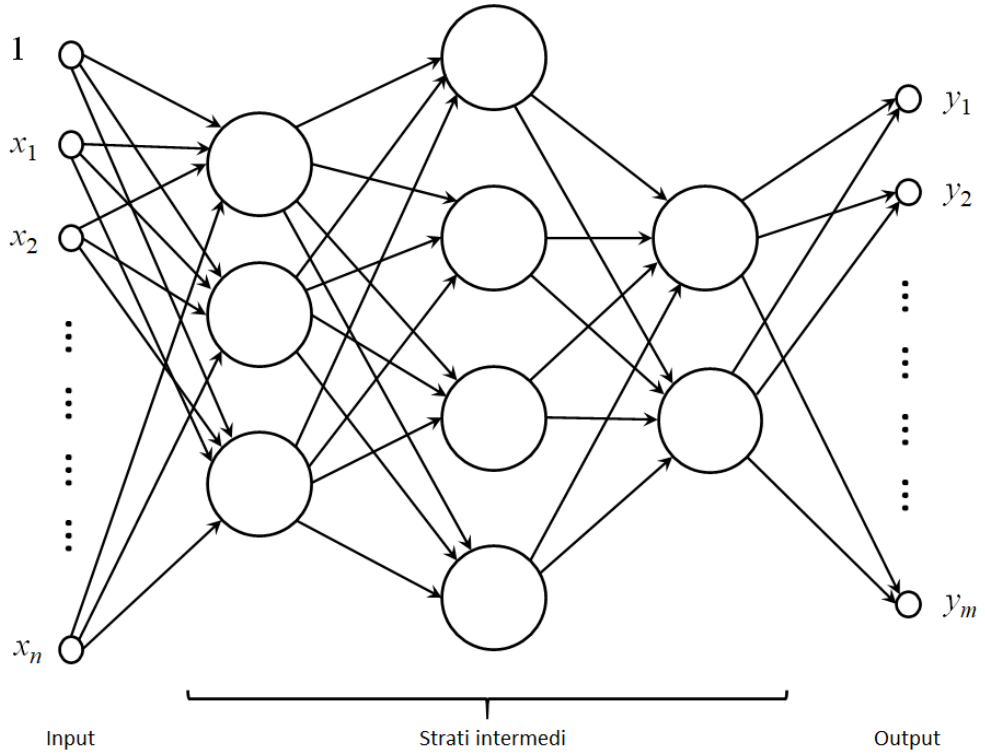


Figura 3 - Una rete neurale artificiale multistrato[38]

Negli anni successivi sono stati definiti numerosi modelli di reti neurali, sperimentando diverse funzioni di aggregazione degli input, diverse funzioni di trasferimento (che calcolano l'output a partire dal valore aggregato degli input), diverse topologie (cioè schemi di connessione tra neuroni), e diverse varianti di algoritmi di apprendimento. Per molto tempo, però, sia il numero totale di neuroni presenti nelle reti che il numero di strati intermedi - cioè compresi tra i neuroni di input e quelli di output - sono stati piuttosto limitati, a causa della ridotta potenza di calcolo dei computer: addestrare una rete neurale dotata di molti strati, e quindi di molti pesi da modificare, risultava proibitivo a causa degli elevati tempi di calcolo. Un altro problema è che per addestrare reti dotate di molti pesi servono tanti esempi di cui si sa qual è l'output desiderato. Infine è da calcolare il dimensionamento della rete: quanti strati intermedi usare? Quanti neuroni mettere in ciascuno strato?

Tutti questi problemi sono stati risolti alla radice grazie alla combinazione di due fattori: l'aumento esponenziale della potenza di calcolo a cui abbiamo assistito nel corso degli ultimi anni (Legge di Moore), e la disponibilità dei cosiddetti *big data*, che forniscono una gran quantità di esempi per addestrare le reti neurali. Negli ultimi anni è così esploso il **deep learning**, con la proposta di diverse tipologie di reti neurali profonde (*deep neural networks*), tra cui spiccano le *deep belief networks*, le *recurrent neural networks* (RNN) e le *convolutional neural networks* (CNN). Per questi modelli di reti sono stati definiti diversi algoritmi di apprendimento supervisionato, semi-supervisionato e non supervisionato, da usare a seconda della quantità di esempi di cui si conosce l'output desiderato (rispettivamente tutti, pochi, nessuno).

[38] Fonte: figura autoprodotta.

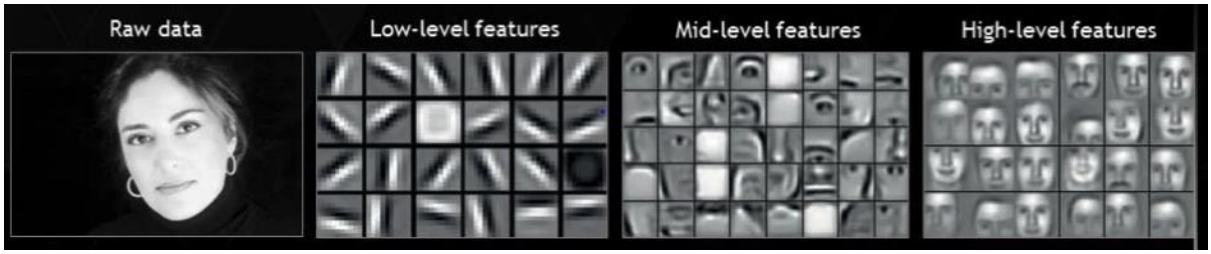


Figura 4 - Dall'immagine di input alle high-level features[39]

Le reti neurali profonde sono state applicate anzitutto - e con molto successo - al riconoscimento delle immagini, grazie a una caratteristica che le rende incredibilmente utili nelle applicazioni pratiche: come si può vedere in Figura 4, mano a mano che ci si sposta dagli strati vicini ai neuroni di input verso quelli vicini ai neuroni di output, i neuroni riconoscono caratteristiche sempre più astratte dell'immagine fornita in input. I primi strati si occupano quindi di riconoscere piccoli segmenti, pezzi di curve o bordi in cui cambia il colore dell'immagine; gli strati successivi usano queste feature per riconoscere caratteristiche più complesse, e così via fino ad arrivare ai neuroni di output, che riconoscono interi oggetti contenuti nell'immagine, consentendo di interpretare il contenuto dell'immagine e il suo contesto.

Questa capacità di astrazione comincia a essere utilizzata anche nell'ambito sicurezza[40], in particolare per l'individuazione di malware, di botnet, di schemi di attacco negli intrusion detection systems (IDS), di attacchi di tipo *drive-by download*, e di spam. Le reti neurali profonde sono inoltre utilizzate nell'autenticazione degli utenti, per capire ad esempio se i caratteri inseriti sono stati digitati da un essere umano.

Concludiamo questa sezione sui modelli connessionisti menzionando il fatto che esiste un intero ramo di ricerca dedicato allo studio di modelli di reti neurali artificiali che cercano di simulare, in maniera più o meno fedele, il funzionamento delle reti di neuroni biologiche, tra cui menzioniamo in particolare il modello *leaky integrate-and-fire* e quello di *Hodgkin–Huxley*. Questi modelli, detti anche **spiking neural networks**, sono utili perché da un lato consentono di investigare e capire come avviene l'apprendimento nelle reti neurali biologiche e dall'altro lato consentono di simulare - e quindi di capire - come avvengono certi malfunzionamenti delle reti di neuroni presenti nel nostro cervello, ad esempio in presenza di attacchi epilettici[41]. Lo svantaggio delle spiking neural networks, soprattutto quelle che simulano in maniera più fedele il funzionamento dei neuroni biologici, è che richiedono una notevole potenza di calcolo.

[39] Fonte: NVIDIA https://www.slideshare.net/NVIDIA/gpuaccelerated-deep-learning-for-cudnn-v2.
[40] Daniel S. Berman, Anna L. Buczak, Jeffrey S. Chavis and Cherita L. Corbett. A Survey of Deep Learning Methods for Cyber Security. *Information* 10(4), 122, 2019. DOI: 10.3390/info10040122.
[41] Rich, S., Hutt, A., Skinner, F.K. *et al.* Neurostimulation stabilizes spiking neural networks by disrupting seizure-like oscillatory transitions. *Sci Rep* **10,** 15408 (2020). DOI: 10.1038/s41598-020-72335-6.

7.4.2 L'approccio evoluzionista

L'approccio evoluzionista parte dal presupposto che le soluzioni a problemi difficili possono essere così complicate da non poter essere concepite, o anche solo realizzate nei dettagli, da esseri umani. Perché allora non mettere in competizione tra loro diverse soluzioni possibili, o *ammissibili*, come avviene in Natura, in modo che vinca la migliore (o meglio, una delle migliori)? Inoltre, anziché avere semplicemente un insieme di soluzioni statiche, che competono per vincere, perché non fare in modo che queste soluzioni *evolvano* nel tempo, adattandosi al problema che devono risolvere?

Sfruttando i principi sottostanti alla teoria dell'evoluzione di Darwin, gli **algoritmi evolutivi** definiscono alcune operazioni per far evolvere gli elementi di una popolazione di soluzioni ammissibili, a ciascuno dei quali viene assegnato un *valore di fitness*, che misura la bontà della soluzione rappresentata. Si pensi ad esempio a un problema di ottimizzazione: più vicina all'ottimo è la soluzione rappresentata dall'elemento, maggiore sarà il suo valore di fitness. Così, nel caso degli **algoritmi genetici**, ogni soluzione ammissibile è rappresentata da una sequenza di bit che rappresentano geni e alleli (forme alternative con cui può presentarsi un gene); vengono poi definiti degli operatori di *crossover*, che a partire da due elementi ne producono uno nuovo, contenente parti del codice genetico di entrambi i genitori; al codice genetico dell'elemento prodotto, infine, si possono applicare delle mutazioni casuali.

Questi principi di base si possono applicare a diverse tipologie di elementi (spesso chiamati *individui*). Nel caso della **programmazione genetica**, ad esempio, viene fatta evolvere una popolazione di *programmi per computer*, costituiti da espressioni o intere funzioni che possono essere valutate da un processore virtuale appositamente definito. Esistono diverse proposte di algoritmi per far evolvere soluzioni ammissibili di problemi, ispirati al funzionamento di diversi fenomeni naturali, quali il volo di stormi di uccelli, la ricerca di cibo da parte di formiche o di api e le tecniche di caccia dei lupi. In tutti i casi, la speranza è che ad ogni generazione compaia nella popolazione un elemento avente valore di fitness maggiore degli elementi delle generazioni precedenti, e quindi più adatto a risolvere il problema considerato.

Nel corso degli anni, gli algoritmi evolutivi sono stati applicati a diversi tipi di problemi. Ad esempio, gli algoritmi genetici sono una componente importante per la generazione di sistemi esperti dell'applicativo FuzzyWorld[42], realizzato dal prof. Lorenzo Schiavina dell'Università Cattolica di Brescia già agli inizi degli anni '90. FuzzyWorld consente di generare una versione ottimizzata del motore inferenziale di un sistema esperto, a partire da dati strutturati in *antecedenti* e *conseguenti*. In questo modo è stato realizzato un sistema esperto che, partendo dal segnale misurato da un sensore in presenza di un mix di concentrazioni di vari inquinanti, è in grado di valutare la concentrazione di uno specifico inquinante che partecipa alla formazione del mix, ad esempio l'ammoniaca[43].

[42] http://www.edor.it/index.php?option=com_content&view=article&id=30&Itemid=65.
[43] Maria Chiesa et al. *Development of low-cost ammonia gas sensors and data analysis algorithms to implement a monitoring grid of urban environmental pollutants*. Journal of Environmental Monitoring. 2012, #14, pagine 1565–1575. doi: https://doi.org/10.1039/C2EM30102D.

Nel caso citato era necessario ottimizzare il numero di insiemi fuzzy di ogni variabile letterale utilizzata nel sistema esperto al fine di migliorare la precisione dei risultati; l'utilizzo degli algoritmi genetici ha consentito di ottenere risultati pressoché impossibili da ottenere con altri strumenti. Un altro esempio di applicazione delle tecniche evolutive, particolarmente interessante, è la cosiddetta **neuroevoluzione**[44], cioè l'evoluzione di reti neurali artificiali. Seguendo i principi della tecnica NEAT (*neuro-evolution of augmenting topologies*), le strutture neurali ottenute tramite algoritmi evolutivi si possono combinare tra loro - come se fossero dei mattoncini elementari non divisibili - usando gli stessi meccanismi di crossover e mutazione, creando così delle strutture a un livello di astrazione più elevato. La sfida è quella di creare reti neurali artificiali che possano un giorno competere con quelle naturali e con una in particolare: il cervello umano.

7.4.3 Altri approcci allo sviluppo dell'IA

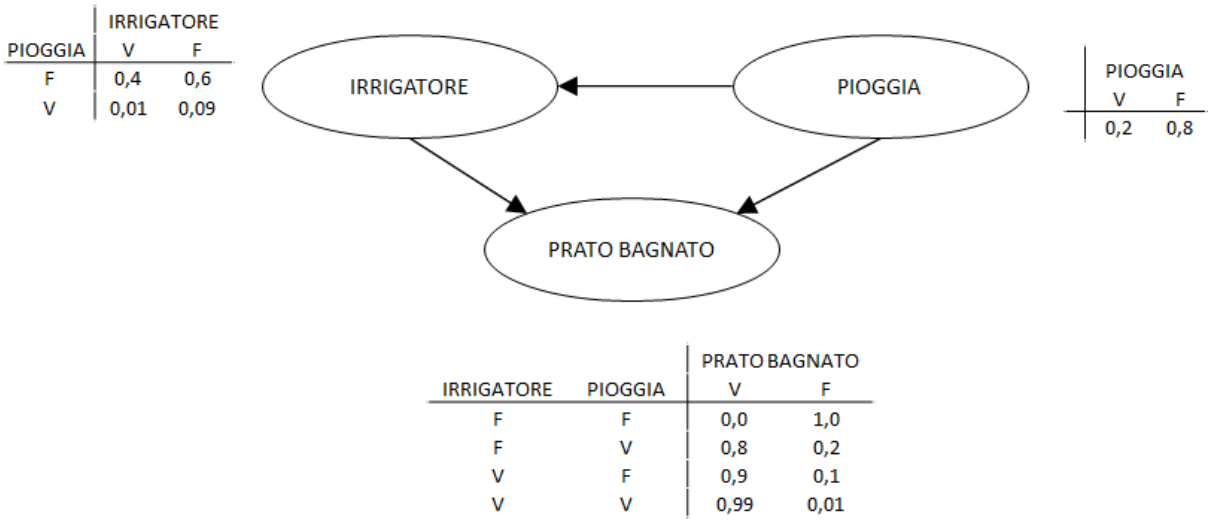
Esistono numerosi altri approcci all'IA, a cui purtroppo non possiamo dedicare lo spazio che meriterebbero. In questo paragrafo menzioniamo solamente tre approcci, che si sono dimostrati molto efficaci in alcune applicazioni pratiche.

Il primo approccio è quello del cosiddetto **automated reasoning**, o **ragionamento deduttivo**. Storicamente è stato il primo approccio all'IA, nel tentativo di simulare il ragionamento deduttivo, basato su evidenze, tipico del cervello umano (considerato da un punto di vista molto astratto, indipendente dal fatto che il cervello sia costituito da una vasta rete di neuroni e altri elementi fisiologici). I sistemi di ragionamento usati nell'IA sono solitamente basati su sistemi logico-deduttivi che partono da un certo numero di assunzioni e - applicando regole di deduzione logica, basate ad esempio sul *modus ponens* e sul *modus tollens* - aggiungono nuovi fatti, o nuove conoscenze, a un database esistente. Sistemi di questo tipo si scontrano tipicamente con problemi di complessità, legati alla dimensione del database e al numero di possibili modi in cui si possono applicare le regole di deduzione logica ai fatti noti. Tuttavia, nel corso degli anni sono state sviluppate numerose tecniche algoritmiche interessanti, sia per *rappresentare la conoscenza* (**knowledge representation**), sia per gestirla durante il ragionamento, utilizzando ad esempio la *logica di default* per gestire fatti come "normalmente gli uccelli volano" (a cui poi andranno aggiunte le eccezioni), oppure la *logica intuizionista* per gestire fatti che non si sa ancora se sono veri o falsi, oppure ancora la *logica fuzzy*, per gestire varianti imprecise di fatti noti, come ad esempio "Carlo è *molto* alto ed è *più simpatico* di Andrea". I sistemi di ragionamento deduttivo sono stati usati anche per costruire i cosiddetti *sistemi esperti*, che hanno riscosso notevole successo - soprattutto negli anni passati - nel risolvere problemi in cui tipicamente è necessario l'intervento di un esperto: diagnosi mediche, prospezioni geologiche, ecc.

Un secondo approccio è costituito dalle **reti bayesiane**, a volte chiamate anche *belief networks.* Questo tipo di modelli si presta molto bene a una rappresentazione grafica, sotto forma di *grafo orientato* (si veda la Figura 7), in cui i fatti sono associati ai nodi del grafo e una freccia tra due nodi - etichettata con un valore di probabilità - rappresenta una possibile

[44] Kenneth O. Stanley et al. Designing neural networks through neuroevolution. *Nature Machine Intelligence*. 2019, #1, pagine 24–35. DOI: https://doi.org/10.1038/s42256-018-0006-z.

relazione causale (probabilistica, condizionata) tra i due fatti. Associando un valore di probabilità ai nodi che non hanno frecce entranti, che sono quindi indipendenti da altri fatti, seguendo le frecce e applicando il *teorema di Bayes*, è possibile calcolare la probabilità degli altri fatti menzionati nel modello. In teoria le reti bayesiane sono un modello *universale* (vedi paragrafo 7.1), e consentono quindi di risolvere qualsiasi problema di apprendimento automatico; in pratica, sono modelli adatti a gestire conoscenze incerte dei fatti e a fare ragionamenti di tipo probabilistico. Esempi sono le applicazioni per le diagnosi mediche, in cui, a partire dai sintomi osservati nel paziente, si deducono le possibili patologie con un determinato grado di probabilità, oppure - per analogia, riferito a macchine anziché alle persone - i sistemi per la diagnosi dei guasti.



PIOGGIA	IRRIGATORE V	IRRIGATORE F
F	0,4	0,6
V	0,01	0,09

PIOGGIA V	PIOGGIA F
0,2	0,8

IRRIGATORE	PIOGGIA	PRATO BAGNATO V	PRATO BAGNATO F
F	F	0,0	1,0
F	V	0,8	0,2
V	F	0,9	0,1
V	V	0,99	0,01

Figura 7 - Un esempio di rete Bayesiana[45]

Un terzo approccio è quello delle **support vector machines (SVM)**, una classe di modelli di apprendimento supervisionato molto potente e molto usato nel machine learning. Consentono di risolvere problemi di classificazione e di regressione. I dati su cui lavorano le SVM sono dei vettori, le cui componenti costituiscono le caratteristiche (*features*) pertinenti con il problema trattato; a ogni vettore è associata un'etichetta che dice a quale classe appartiene il vettore (per semplicità, si pensi a due sole classi, come ad esempio *buono* o *cattivo*). Rappresentati questi vettori come punti in uno spazio *n*-dimensionale, le SVM, nella loro versione di base, cercano un iperpiano (cioè una retta in uno spazio bidimensionale, un piano in uno spazio tridimensionale, ecc.) che separi i punti appartenenti alle due classi, nel senso che tutti i punti di una classe stanno da una parte e tutti i punti dell'altra classe stanno dall'altra parte. Inoltre, tra tutti gli iperpiani possibili, ne viene individuato uno per cui la distanza dall'iperpiano a qualsiasi punto delle due classi è la massima possibile. Per consentire alle SVM di risolvere problemi che coinvolgono più di due classi e problemi più difficili di quelli *linearmente separabili* (per i quali cioè esiste un iperpiano che separa le classi), sono state proposte diverse varianti di SVM. In particolare, l'uso di *funzioni kernel* e del cosiddetto **kernel trick** consente di applicare una funzione appropriata allo spazio dei vettori di input, in modo da trasformare un problema di classificazione non-lineare (per il quale occorrerebbe una funzione

[45] https://commons.wikimedia.org/wiki/File:SimpleBayesNet.svg (tradotta).

più complicata di un iperpiano per separare le classi) in un problema di classificazione lineare (si veda la Figura 8).

In teoria, anche le SVM costituiscono un modello di apprendimento *universale* (vedi paragrafo 7.1). Tuttavia, soprattutto nel caso della classificazione non-lineare, l'elevato numero di parametri da regolare affinché l'algoritmo di apprendimento produca dei buoni risultati, ne rende l'applicazione più difficile rispetto ad altri modelli.

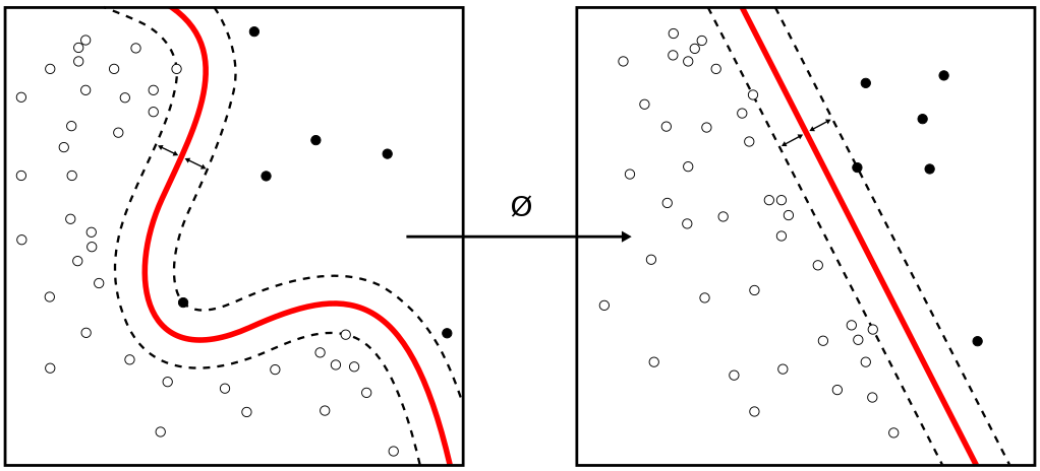


Figura 8 - SVM: trasformazione da un problema di classificazione da non-lineare a lineare[46]

[46] Attribuzione: Alisneaky, svg version by User:Zirguezi, CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0), via Wikimedia Commons. URL: https://commons.wikimedia.org/wiki/File:Kernel_Machine.svg.

8 I progetti di IA

Un progetto IA ha caratteristiche proprie peculiari e altre non difformi da un qualsiasi progetto di trasformazione digitale.

Un progetto di introduzione di una soluzione IA in un'organizzazione ha alcune caratteristiche comuni a tutti i progetti di trasformazione digitale e ne ha altre peculiari.

Una soluzione basata sull'IA deve essere progettata fin dall'inizio in maniera coerente con gli obiettivi e con gli strumenti utilizzati, oltre che con i vincoli legali e regolamentari. Essendo possibile impiegare moltissimi strumenti, occorre scegliere fin dal principio quelli giusti.

8.1 I dati

Un sistema di IA necessita di dati di qualità per poter fornire i risultati desiderati. La maggior parte delle organizzazioni non è in grado di misurare lo stato di salute dei propri dati e il 62% ha affermato di non fidarsi dei propri dati[47].

È quindi necessario stabilire quali sono le caratteristiche per cui i dati possono essere considerati di qualità. I dati devono essere:

- accurati e privi di errori (cioè che siano stati sottoposti ad un esistente / predeterminato insieme di regole e controlli);
- completi (cioè che ragionevolmente rispecchino le possibili alternative e variazioni);
- aggiornati (ossia che siano recenti ed elaborati il prima possibile per sfruttare pienamente le capacità della IA e per rispondere tempestivamente a cambiamenti interni o esterni);
- conformi rispetto alle regole di privacy e sicurezza del settore (ossia raccolti, in forza di corretta base giuridica, nel rispetto dei principi di trasparenza e correttezza e del principio di minimizzazione e, infine, che siano protetti con misure di sicurezza allo stato dell'arte).

Un modello basato su dati, analisi e IA richiede molta coerenza nella scelta delle fonti dei dati da raccogliere, nella classificazione dei dati e nella loro modalità di conservazione. La frammentazione tipica di molte organizzazioni (database, software, processi a silos) è ben noto che depotenzia l'introduzione dell'IA. Qualora la centralizzazione fosse impossibile nel breve, l'organizzazione dovrebbe provvedere a disporre di linee guida esplicite e standard relativi a legalità e autorizzazione di raccolta, catalogazione, ubicazione, uso e sicurezza dei dati.

[47] https://www.talend.com/.

8.2 Le caratteristiche peculiari di un progetto IA

Per i progetti di adozione dell'IA vanno applicati alcuni principi al fine di poterne trarre i benefici auspicati[48].

- L'IA dovrebbe essere utilizzata su una **base integrata** di dati, analisi e software.
- Vanno aggiornate le **valutazioni del rischio**, includendo i nuovi rischi e le relative misure di controllo. Le valutazioni del rischio devono permettere di individuare, nei limiti delle competenze e budget disponibili, le misure necessarie per proteggere il sistema e i suoi utilizzatori da attacchi e frodi. Le valutazioni del rischio vanno periodicamente riesaminate per identificare eventuali necessità di aggiornamento.
- Può essere estremamente difficile integrare i modelli di IA nell'architettura tecnologica complessiva esistente[49]. È quindi fondamentale, per evitare l'insorgere di problemi nelle fasi successive, **verificare fin dall'inizio se la soluzione IA possa essere sviluppata e integrata nell'ambiente IT esistente.**
- Quando si sceglie una soluzione basata su tecnologie IA è importante comprendere come vengono istruiti gli algoritmi, **come vengono aggiornati e come sono validati i dati**. Tutte queste informazioni sono necessarie per poter identificare al meglio la soluzione più idonea in base allo specifico contesto.
- Relativamente alle **competenze,** è necessario distinguere tra quelle necessarie al progetto e quelle delle persone utilizzatrici delle soluzioni IA. Per il primo caso bisogna acquisire (attraverso formazione o acquisizione sul mercato) competenze relative al contesto, alla tecnologia specifica e alla sicurezza informatica. Le figure tipiche da prevedere in un progetto IA sono[50]: responsabile di progetto (profondo conoscitore dell'attività primaria dell'organizzazione), responsabile di prodotto IA, esperto di algoritmi, data scientist, data engineer, data architect, visualization specialist, user interface designer, e business analyst. Per quanto riguarda gli utilizzatori delle soluzioni IA, è necessario che siano in grado di consultare gli strumenti di IA e prendere decisioni adeguate sulla base delle loro raccomandazioni. L'ingaggio di tutte le persone, portate a sperimentare i benefici dell'IA sulla propria attività, è fondamentale per il successo delle iniziative[51].
- La **governance delle risorse digitali** deve essere multidisciplinare perché richiede, oltre che competenze tecniche e digitali, anche conoscenze normative (GDPR in primis e altre normative specifiche nazionali e comunitarie), di analisi dei dati (per identificare eventuali algorithmic bias, approfonditi nel capitolo 33.1.3) e di sicurezza informatica. L'intelligenza artificiale richiede approfondimenti sulle sfide legali, etiche e di sicurezza, inclusa un'attenta valutazione di quali dati possono o non possono essere conservati.

I progetti IA hanno poi caratteristiche comuni a tutti i progetti in ambito informatico.

[48] https://hbr.org/2020/01/competing-in-the-age-of-ai.
[49] https://hbr.org/2020/06/the-dumb-reason-your-ai-project-will-fail.
[50] https://www.computersciencedegreehub.com/faq/skills-job-artificial-intelligence/
[51] https://hbr.org/2020/07/a-better-way-to-onboard-ai.

- Le soluzioni di IA sono strumenti **al servizio della strategia** e non obiettivi fini a sé stessi; vanno quindi analizzati gli obiettivi e i problemi esistenti, individuando quelli che possono beneficiare dell'implementazione di soluzioni di IA attualmente disponibili.
- È opportuno definire un portafoglio di progetti che comprenda dei "**quick win**", in modo da acquisire gradualmente esperienza sulla tecnologia e costruire il consenso all'interno dell'organizzazione, necessario per portare avanti le iniziative a più ampio respiro, strategiche e trasformative.
- La **scalabilità** delle soluzioni e del resto dell'infrastruttura IT è un prerequisito per poterne sfruttare le potenzialità. Se i dati da caricare a sistema sono ingenti, bisogna evitare i colli di bottiglia, creare un ambiente affidabile e progettare architetture di elaborazione e archiviazione che non creino problemi di velocità e latenza.
- Così come le organizzazioni sono soggette a continui cambiamenti, così la soluzione deve essere sufficientemente **flessibile** per adattarsi alle nuove situazioni e senza che ne vengano compromesse le funzionalità. L'approccio alla manutenzione deve considerare la possibilità di dover riconfigurare il sistema e includere un meccanismo di sincronizzazione dei dati rapida e senza problemi.
- Per assicurare la flessibilità, può essere utile sfruttare gli **approcci "agile"** perché permettono una maggiore capacità di sperimentazione e di apportare cambiamenti al progetto, inclusi quelli correttivi.

8.3 Il ciclo di sviluppo di una soluzione IA

L'approccio che negli ultimi anni sta dando i migliori risultati è quello "agile". Si lavora a rilasci continui in una logica di MVP (minimum viable product) sapendo che i primi rilasci - e anche quelli intermedi - non saranno perfetti. È meglio rilasciare una soluzione "non perfetta" piuttosto che aspettare mesi e mesi di sviluppo per avere un algoritmo quasi perfetto, ma rilasciato troppo tardi o che non risponde più ai requisiti che nel frattempo potrebbero essere cambiati.

È estremamente importante che vengano forniti riscontri sull'utilizzo dei primi MVP in modo da ottimizzare e migliorare l'algoritmo a ogni successivo rilascio. Per questo si parla di un processo iterativo ed incrementale.

Figura 9 - Processo di realizzazione di progetti IA[52]

Le fasi del ciclo di vita di sviluppo di un caso d'uso di IA sono schematizzate in Figura 9. La durata di ciascuna fase è indicativa e basata sulle esperienze degli autori. Essa può variare a seconda del contesto e delle caratteristiche della soluzione di IA, ma è sempre relativamente breve (uno o due mesi).

1. **Comprensione del business:** in questa fase, che dura all'incirca un mese, come dettato dall'esperienza, il gruppo di lavoro si concentra sulla comprensione e l'approfondimento del requisito di business, identificando con precisione l'obiettivo e definendone il perimetro.
   È fondamentale identificare fin da subito le metriche di valutazione della bontà dei risultati, per accordarsi in modo chiaro sui criteri di valutazione della qualità del lavoro svolto e delineare un piano di alto livello del progetto.
   In questa fase si avvia l'aggiornamento della valutazione del rischio.

2. **Comprensione dei dati di input**: il gruppo di lavoro identifica i dati necessari per addestrare o verificare l'applicazione, ne verifica la qualità, la facilità di accessibilità e utilizzabilità, e ne identifica eventuali criticità, anche per quanto riguarda il trattamento di dati personali. I dati possono essere già tutti a disposizione o necessitano un'integrazione all'interno dei sistemi dell'organizzazione. In una prima fase si producono risultati intermedi o preliminari che possono essere di ulteriore spunto di discussione e approfondimento per migliorare i dati di input. Le organizzazioni che, al loro interno, hanno gruppi che si occupano di data governance e qualità dei dati sono facilitate nel generare il valore desiderato nel minor tempo possibile. Questa fase di comprensione dei dati dovrebbe durare non più di un mese.

[52] Figura autoprodotta.

3. **Preparazione dei dati**: le persone con profili di data engineer e data scientist hanno le responsabilità principali di questa fase. Le attività includono la selezione e pulizia dei dataset appropriati, l'integrazione di vari dataset precedentemente uniformati, la costruzione delle variabili necessarie per l'algoritmo e la progettazione e preparazione del dataset finale (*master table*) sul quale addestrare (per gli algoritmi che lo prevedono) o verificare la soluzione di IA.

4. **Sviluppo del modello:** una volta pronta la master table, si passa alla fase di sviluppo del modello matematico. La prima attività è quella di selezionare le tecniche di modellizzazione più rilevanti per il caso d'uso che si sta affrontando. Se il tempo lo consente è bene provare almeno due o tre diverse tecniche, in modo da confrontare i risultati preliminari e decidere quale tecnica si addice meglio a rispondere al requisito di business. Successivamente si passa all'ottimizzazione dei singoli componenti e del modello in generale. Esistono diverse tecniche di ottimizzazione, ognuna con i propri punti di forza e di debolezza. L'ottimizzazione non è mai un processo che si fa una volta sola ma - a seconda del problema - è necessario utilizzare più tecniche e confrontarne i benefici. La fase di sviluppo del modello si conclude con la valutazione dello stesso da un punto di vista puramente statistico. Se, per esempio, l'obiettivo è quello di ottenere un modello predittivo, possono essere usate tecniche come la *matrice di confusione* o la *curva ROC*.

5. **Valutazione**: la valutazione del modello avviene quando gli utilizzatori iniziano a usarlo per le loro attività. Solitamente non si mette in produzione il modello senza dedicare almeno un mese ai test in grado di verificarne l'efficacia. Durante questa fase vengono raccolte informazioni utili per migliorare l'impostazione del modello stesso. In questa fase si individua anche il sistema dei controlli di processo e tecnologici per affrontare i rischi.

6. **Integrazione**: una volta che il modello è robusto da un punto di vista matematico e ha ottenuto delle buone valutazioni in fase di test, si passa all'integrazione all'interno dei sistemi informatici dell'organizzazione. In questa fase vanno seguiti i processi e le procedure di gestione dei cambiamenti dell'organizzazione, prestando attenzione alle tempistiche e al sistema dei controlli. Per una maggiore rapidità, si raccomanda di adottare un approccio DevOps, che, insieme all'approccio allo sviluppo agile, consente di avere un algoritmo sempre aggiornato e prestante.

8.4 Integrare l'IA nei processi

Anche se l'IA non è una recente novità (per esempio, negli anni Ottanta, nel Gruppo Montedison, fu sperimentata un'applicazione dedicata alla prevenzione delle malattie delle verze realizzata dal Team IA aziendale per la divisione fitofarmaci), allo stato attuale è ragionevole considerarla una novità "disruptive" per due ragioni fondamentali:

- la difficoltà delle organizzazioni nel comprenderne l'impatto potenziale sui processi;
- la preparazione, spesso carente, delle persone dedicate alla progettazione e realizzazione di soluzioni di IA, in area tecnologica e organizzativa e nell'ambito dei controlli.

Come superare tale situazione che costituisce una barriera alla possibilità di cogliere le innegabili opportunità dell'IA? Potrebbero essere considerate due possibilità:

- percorrere strade già collaudate in passato per la gestione dell'innovazione ricorrendo alla leva formativa e promuovendo un progressivo cambiamento nei profili di riferimento manageriali e tecnologici;
- cercare nuovi approcci di non semplice identificazione.

Probabilmente la soluzione ideale da perseguire potrebbe essere un mix delle possibilità citate, come di fatto è avvenuto sul mercato a fronte di cambiamenti rilevanti come l'avvento dei sistemi informatici real time nel decennio 1965-75, o l'avvento di Internet negli anni Novanta.

È quindi necessario pianificare il patrimonio di competenze e i profili professionali che ormai le organizzazioni non possono più trascurare.

https://www.youtube.com/watch?v=dNZXLH4xeAs

"How to insert AI into business processes and what are the critical factors to consider?"

Una delle principali tecnologie è l'intelligenza artificiale (AI).
Come renderla un vantaggio?
Il video offre alcuni segreti sulle trasformazioni basate sull'IA di successo.

(durata: 10.02 minuti)

8.4.1 Intelligenza artificiale e umana

Risulta interessante evidenziare come, a seconda della tipologia di attività considerata, risulti differente il mix ottimale tra intelligenza artificiale e intelligenza umana che è utile implementare.

A un estremo dello spettro delle attività si collocano quelle che possono essere demandate a una macchina più o meno intelligente (per esempio: adattamento a condizioni differenti, previsione di eventi futuri, ripetizione di azioni, transazioni predefinite); all'altro estremo si possono individuare attività che, allo stadio attuale dello sviluppo tecnologico, devono essere ancora svolte in massima parte dall'uomo (per esempio, quelle che richiedono di esprimere un giudizio, creatività, empatia, leadership).

Tra i due estremi, si incontrano quelle attività in cui intelligenza artificiale e umana si possono complementare al meglio con un'attenta riprogettazione dei processi. Si hanno quindi, da una parte, attività in cui l'IA ha un peso prevalente e fornisce all'uomo "superpoteri" (si pensi all'aumentata capacità di interazione che può dare un traduttore simultaneo o all'amplificazione delle capacità visive resa possibile da un visore intelligente) mentre dall'altra si ha il caso in cui l'uomo affianca l'IA per supervisionarne le decisioni e renderla più efficace (come ad esempio nella spiegazione delle decisioni di un algoritmo o nelle attività di formazione).

Nell'ambito del dibattito sull'etica dell'IA (non oggetto di questo libro), le considerazioni precedenti possono fornire un'utile linea guida su come spiegare il ruolo che l'IA, al di là di scenari ad oggi futuristici, deve poter svolgere al servizio dell'uomo (cosiddetta *augmented intelligence*) e per prevedere quali saranno le attività in cui l'uomo verrà più velocemente sostituito dall'IA.

8.4.2 La *robotic process automation*

La *robotic process automation* (RPA) consiste nell'automatizzazione di mansioni ripetitive svolte da risorse umane a robot software (p.e. l'approvazione di mutui bancari).
Quando trasferiamo le attività effettuate dalle risorse umane a robot software parliamo di RPA: *robotic process automation*.

La prima fase dell'automazione simula l'attività umana: il codice dei robot software replica l'agire dell'addetto umano passo dopo passo. La seconda fase è detta *di semplificazione;* il codice viene riscritto eliminando passaggi superflui e tipicamente umani, in quanto non più necessari. Per esempio, la conferma del salvataggio di un dato, necessaria come controllo per evitare un errore della persona, deve essere rimossa in quanto inutile, i robot software non sbagliano.

La RPA abilita la revisione cioè il ridisegno dei processi aziendali integrando le tecnologie che operano con contenuti digitali.

Perché utilizzare la tecnologia RPA? Per sottrarre la risorsa umana dalle attività ripetitive, soggette ad errore, con prestazioni scarse e limitate alla disponibilità lavorativa. I robot possono funzionare sempre, senza errori, con migliori prestazioni e senza limitazione di fascia oraria.

È per questo che RPA e IA vivono e vivranno sempre più in simbiosi di mutualismo, una simbiosi ove esiste un vantaggio reciproco per l'evoluzione di entrambe le tecnologie. Nasce il termine IPA, *intelligent process automation*, indicando l'integrazione di queste nuove tecnologie.

Proviamo qui sotto ad elencare gli ambiti dove questa simbiosi sta già avvenendo.

- Acquisizione semiautomatica dei processi da automatizzare da parte delle piattaforme tecnologiche di RPA quali la computer vision. La tecnologia IA presente negli strumenti di RPA permette di riconoscere gli elementi di scambio delle informazioni fra l'uomo e i sistemi. Per esempio è possibile automatizzare l'inserimento dati in una maschera di un software perché la tecnologia riconosce i campi di input anche se subisce modifiche quali la dimensione o posizione, la descrizione o il colore.
- Acquisizione dei contenuti di documenti non strutturati che utilizzano IA con tecnologie OCR per trasformare un'immagine, anche di bassa qualità, in stringhe di caratteri.
- Alimentazione di un sistema di machine learning con volumi di dati digitali sempre più crescenti riducendo i tempi di acquisizione dei dati stessi. Le piattaforme di RPA sono in grado di sottomettere radiografie digitali di pazienti sospettati di avere una patologia polmonare a un motore di IA per analizzare il referto e ottenere l'esito della diagnosi associato ad un indice di affidabilità. Si ottengono tempi dimezzati nell'ottenimento della diagnosi e nella gestione di volumi significativi.

Intervista a Darya Majidi, *Founder & CEO di Daxo Group e di Dcare*

Imprenditrice, laurea in informatica, master in economia, TedX speaker, docente, femminista 4.0, autrice dei libri "Donne 4.0", "Sorellanza Digitale e "Connected Sisterhood"; menzionata da D. di Repubblica nel 2020 come una delle 100 donne che cambiano il mondo.
Si è laureata in Scienze dell'informazione con specializzazione in Intelligenza Artificiale, all'Università di Pisa, oltre a conseguire un Master in "Strategia e governance aziendale" al Dipartimento di Economia. A 28 anni ha creato la sua prima start up, Synapsis, spin-off della Scuola Superiore Sant'Anna di Pisa, oggi parte di Dedalus Group, dopo una positiva esperienza di private equity.
Presidente del Gruppo Giovani di Confindustria Livorno (2004-2007) e Vice Presidente di Confindustria Livorno (2008-2009) ha avuto la delega alla ricerca e all'innovazione creando il "Club degli Innovatori" con un focus costante sui giovani e sulle startup.
Assessore alla Semplificazione e allo Sviluppo Economico del Comune di Livorno (2009-2014), ha contribuito a trasformare la sua città in una smart city, creando una fitta rete di aree wifi gratuite, portando la fibra in tutta le città e digitalizzando i processi e servizi principali del Comune, in primis l'Anagrafe e gli Sportelli per le imprese ed i cittadini. Inoltre, ha avviato con la Scuola Superiore Sant'Anna il progetto di "Robotica Educativa" nelle scuole superiori della città.
Founder & CEO di Daxo Group, società di consulenza strategica in Industria 4.0, con focus sulla Digital Transformation attraverso progetti di Iot & AI. Founder e Presidente di Dcare, società del gruppo Dedalus, con focus sull'IoT in Sanità.

Domanda 1. Darya, l'intelligenza artificiale è sempre più intorno a noi. Quale approccio adottare per attuare la trasformazione di un'azienda tradizionale in un'azienda 4.0 che fa sempre più uso di questa tecnologia?

Innanzitutto, è indispensabile aver accesso a conoscenze e competenze di professionisti in grado di affiancare l'azienda in questo processo di cambiamento. Trattasi di un passaggio dallo stato attuale ad un nuovo stato: questo passaggio richiede skills particolari per prevedere, disegnare e governare nuove strategie. Diventa fondamentale e strategico trasformare il *mindset* dei manager e degli imprenditori attraverso lo sviluppo di competenze distintive per l'Industria 4.0 che si basa, oggi più che mai, attraverso i numerosi sensori ed applicazioni IoT, IIoT (Industrial Internet of things) e machine learning su una intelligenza artificiale in continuo divenire.

Secondo me è indispensabile portare avanti la cultura dell'innovazione e del cambiamento nelle aziende attraverso il trasferimento del valore delle tecnologie, leva strategica dello sviluppo delle nostre imprese, mettendo sempre al centro dei processi di trasformazione la

persona, l'uomo. In particolare, ritengo che per un buon funzionamento dell'intelligenza artificiale sia necessario avere un mix di conoscenze che sfociano in 4 competenze approfondite, ovvero competenze strategiche, manageriali, digitali e soft skill.

Dobbiamo essere consapevoli che l'intelligenza artificiale è destinata a ricoprire sempre più un ruolo centrale nello sviluppo del business e nei nostri processi produttivi, grazie al miglioramento della relazione esistente tra questa tecnologia esponenziale e l'intelligenza emotiva: due mondi apparentemente lontani ma in realtà estremamente prossimi ed interconnessi già oggi nel nostro quotidiano e nelle aziende. Pertanto, si devono acquisire gli strumenti conoscitivi per il governo dell'intelligenza artificiale all'interno del contesto aziendale ed al contempo di sviluppare la leadership dei manager attraverso l'impiego della propria intelligenza emotiva. Le neuroscienze ci confermano il ruolo centrale delle emozioni nei nostri processi cognitivi, per questo diventa importante saperle governare.

Domanda 2. Ritieni che l'intelligenza artificiale sia condizionata da chi costruisce gli algoritmi fino a convertirsi in una minaccia, un pericolo per la società?

Mi piace ripetere sempre che è necessario rendere consapevoli tutti che le tecnologie stanno plasmando il nostro futuro e il mondo intero e che non possiamo escludere il 50% della conoscenza mondiale - ovvero quella femminile - dalla creazione del mondo. Ovvero: risulta fondamentale dare consapevolezza tecnologica e digitale alle donne di tutto il mondo.
Esistono ormai numerosi studi che dimostrano che progetti di intelligenza artificiale non includono le donne nella fase di progettazione**.** Ne consegue che rischiamo di creare dei sistemi di intelligenza artificiale non funzionanti correttamente, perché contenenti quello che viene chiamato in gergo "*unconscious bias",* ovvero: chi crea il sistema inserisce dei *bias,* i.e. dati errati, dati mancanti, classificazione errata in modo non consapevole.
L'Intelligenza Artificiale non è neutra: nel momento in cui deleghiamo solo ai maschi la scelta dei dati su cui allenare questi algoritmi e le conoscenze da coinvolgere, non sarà possibile garantire la rappresentazione completa della società e potremmo avere dei risultati distorti e gli algoritmi potrebbero non funzionare correttamente per tutti i "colori" dell'arcobaleno della società.

Domanda 3. **Per cui - facendo una trasposizione ardita e filosofica - mi stai dicendo che, come Feuerbach diceva "L'uomo è ciò che mangia" possiamo dire che l'intelligenza artificiale è "ciò che mangia", ovvero, come si dice nel gergo informatico: "garbage in, garbage out"?**

Nell'intelligenza artificiale è fondamentale che i dati selezionati - per essere inclusi nel training set - siano il più completi possibile e rappresentativi del dominio in studio. Non possono mancare i dati significativi ed è ovviamente fondamentale non dare classificazioni errate o mancanti. Pertanto, chi effettua la scelta dei dati e della loro etichettatura, inserisce, senza volerlo, la propria conoscenza (con i suoi limiti), il proprio modello dei dati e di classificazione nel sistema. Ne deriva che è vitale, per avere un sistema corretto ed ampio, una accurata selezione dei dati del training set e una loro classificazione sapiente, coinvolgendo team, non solo multidisciplinari, ma anche che includano la diversity in tutte le forme. La conoscenza, la rappresentazione e l'astrazione di molti domini variano in base al genere, ne deriva che se gli algoritmi sono creati da soli uomini, non hanno la completezza totale nella conoscenza di quel dominio.

Pensiamo ai sistemi di ML supervised, di fatto, non fanno altro che inserire conoscenze esistenti in sistemi in grado di generalizzare questa conoscenza. I sistemi di ML unsupervised invece cercano nei dati non classificati a priori, pattern per creare dei clusters che hanno dati "simili". Questi sistemi, se non supportati da conoscenze preliminari del dominio da esperti, spesso trovano classificazioni e conoscenze di base già note agli esperti. Sempre più, usando gli attuali sistemi di ML supervised e unsupervised ci accorgiamo delle loro lacune, limitazioni e *man-oriented* che possono avere impatti irreversibili sulla società con implicazioni etiche e biologiche da gestire (es. diagnostica e elaborazione algoritmi per medicinali senza considerare etnie, diversità fisiologiche tra uomo e donna, ecc.).

A mio parere un'altra grande area dell'intelligenza artificiale che è destinata a svilupparsi sempre più è il *symbolic learning* (SL) che si basa sul concetto di estrarre la conoscenza di esperti umani e di formalizzarla in basi di conoscenza (*knowledge base*) e di utilizzare delle regole, per trasformare i dati di input in conoscenza in output (motore inferenziale). La conoscenza umana deve essere quindi codificata. Ma quale conoscenza? Non solo la conoscenza pubblica e dichiarativa di un dominio che può essere estratta dai libri e dalle pubblicazioni, attraverso motori di ricerca anche basati su ML, ma soprattutto, la conoscenza cosiddetta procedurale e privata di vari esperti nei propri settori

Nei sistemi di supporto e di esperti basati sul SL, tali regole possono essere codificate nei *knowledge base* per essere a disposizione di tutti. Tuttavia, ancora una volta è doveroso evidenziare come, nel campo tecnologico e nella progettazione degli algoritmi la conoscenza umana è privata di quel 50% della mente della popolazione femminile che, purtroppo, è spesso esclusa dalla codifica dei knowledge base dei sistemi di intelligenza artificiale. Spesso mi domando e pongo la stessa domanda a chi mi ascolta o legge: "Ce lo possiamo permettere di "perdere" metà della conoscenza universale?". A oggi solo il 13% delle risorse tecniche coinvolte nei progetti di IA sono donne ed è ovvio che questo sta creando dei sistemi incompleti e ricchi di bias che, paradossalmente, faranno da amplificatori a questi bias.

È ben noto, ad esempio, il caso di un sistema di AI di selezione di risorse umane di Amazon: i tecnici avevano creato un sistema per la scelta dei curricula di persone talentuose da coinvolgere nel proprio organico ed hanno notato come il sistema non selezionasse mai le donne. Dopo alcune verifiche hanno capito che nel training set non avevano inserito una buona rappresentanza di curricula femminili.

Per questo diventa fondamentale coinvolgere le donne in tutte le fasi e in tutte le discipline dell'intelligenza artificiale, dalla fase di selezione degli obiettivi del sistema di intelligenza artificiale, alla selezione dei dati e delle classificazioni, alla selezione della conoscenza da codificare nei knowledge base, ecc.

Ora con l'arrivo della quarta rivoluzione industriale, grazie a Internet e al cloud, riusciamo a creare e ad accedere a grandi quantità di dati. Le potenze delle macchine stanno crescendo in modo esponenziale e fondamentale sarà garantire un'intelligenza artificiale che continua a utilizzare, in modo interoperabile e interdisciplinare, i sistemi simbolici e "connessionisti" in modo da supportare la crescita della conoscenza umana attraverso lo sviluppo dei cosiddetti sistemi di intelligenza artificiale ibridi, in grado di far confluire le diverse anime dell'intelligenza artificiale. Dobbiamo unire la potenza di calcolo delle macchine con la conoscenza codificata di esperti per arrivare ad una intelligenza "aumentata". E non possiamo, pertanto, non coinvolgere le donne, i loro talenti, le loro conoscenze, in questo nuovo modello.

Domanda 4. Ma che futuro ci attende secondo te: ci troveremo ad interagire sempre più con un'intelligenza artificiale dominante oppure pensi che si convertirà in una leva potente, un'alleata preziosa per l'umanità?

Io sono innamorata dell'intelligenza artificiale e sono convinta che possa essere positiva e messa a servizio dell'uomo: si aprono nuovi scenari di ricerca davvero appassionanti, dove far collaborare macchine e umani per unire l'immaginazione e la creatività con la velocità di calcolo e di accesso a dati e conoscenze disponibili fino a pochi anni fa impensabili. I sistemi cognitivi del futuro dovranno beneficiare di tutte le tecniche dell'intelligenza artificiale e, ancora una volta, quasi come se fosse un mantra, voglio sottolineare l'importanza del coinvolgimento delle conoscenze delle donne, se davvero vogliamo che le macchine siano a servizio dell'intera umanità. Pertanto, anche l'intelligenza artificiale dovrà dimostrarsi "sostenibile", "obiettiva" ed inclusiva, coinvolgendo le donne per creare un mondo più equo, più giusto, più inclusivo.

- TERZA PARTE: USI DELL'IA

9 I tanti usi dell'intelligenza artificiale

L'utilizzo dell'intelligenza artificiale è oggi un vantaggio significativo e comprovato. Per le aziende commerciali, il suo uso, anche grazie a nuove applicazioni e prestazioni sempre più elevate, sarà sempre più necessario per assicurarsi il vantaggio competitivo.

In questa pubblicazione dedicata alla sicurezza è dato ampio spazio agli ambiti di applicazione dell'intelligenza artificiale perché è necessario comprenderli per identificare i rischi da affrontare e i requisiti legali e regolamentari da rispettare.

Le organizzazioni che utilizzano le tecnologie IA possono sfruttare i dati a cui hanno accesso sia per estrarre nuove informazioni, sia per automatizzare i processi o renderli più efficienti. In alcuni casi queste applicazioni sostituiscono gli esseri umani nelle attività ripetitive, in altri casi li affiancano nei loro compiti, in altri casi ancora abilitano processi e funzioni completamente nuovi.

"*Ora immagina cosa succederebbe quando un'azienda basata sull'intelligenza artificiale competesse con un'azienda tradizionale, servendo gli stessi clienti con una proposta di valore simile (o migliore) e un modello operativo molto più scalabile. Potremmo chiamare questo tipo di confronto* collisione*. Le aziende costruite su un nucleo digitale possono sopraffare le organizzazioni tradizionali*"[53].

Le innovazioni che portano discontinuità (*disruptive*) non sono l'emergere di una nuova tecnologia o di un nuovo modello di business: sono la nascita di un nuovo tipo di organizzazione completamente diversa, tale da cambiare radicalmente il proprio settore di riferimento e rimodellare la natura del vantaggio competitivo.

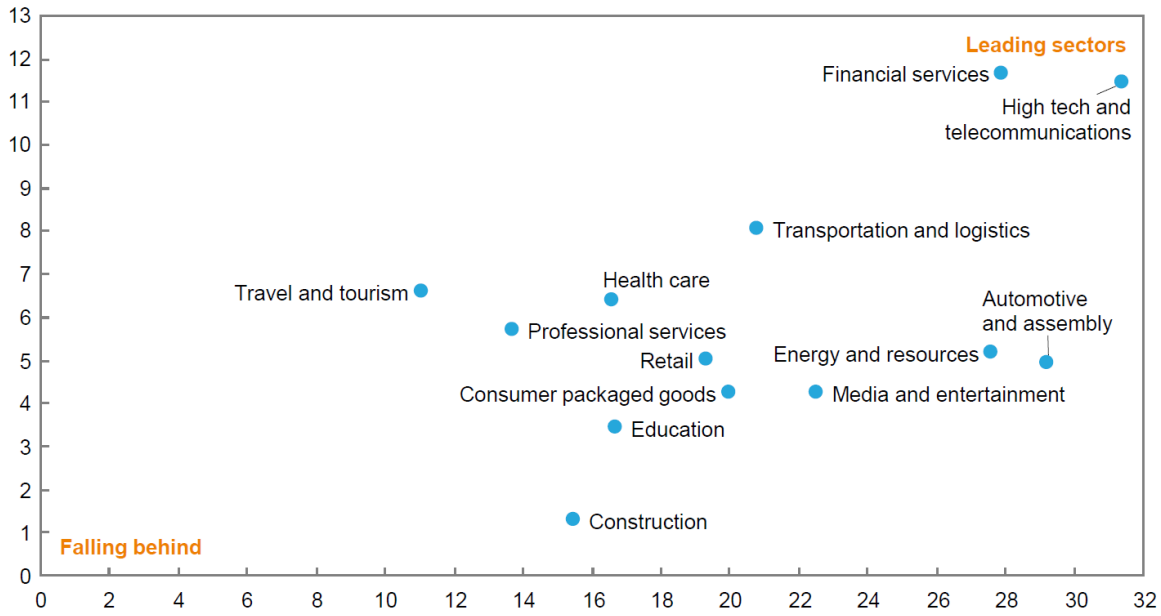
Potrebbe volerci un po' di tempo prima che i modelli operativi basati sull'intelligenza artificiale generino valore economico.

In alcuni settori l'IA sarà adottata in modo più spinto che in altri, come si deduce dal grafico seguente[54]. Questo non va interpretato come incompatibilità fra l'uso dell'IA e il proprio settore o una presunta superiorità del modello tradizionale.

[53] https://hbr.org/2020/07/a-better-way-to-onboard-ai.

[54] Si veda anche la figura 2 dell'articolo "Global AI Survey: AI proves its worth, but few scale impact", disponibile su: https://www.mckinsey.com/featured-insights/artificial-intelligence/global-ai-survey-ai-proves-its-worth-but-few-scale-impact#.

Future AI demand trajectory[1]
Average estimated % change in AI spending, next 3 years, weighted by firm size[2]



1 Based on the midpoint of the range selected by the survey respondent.
2 Results are weighted by firm size. See Appendix B for an explanation of the weighting methodology.

SOURCE: McKinsey Global Institute AI adoption and use survey; McKinsey Global Institute analysis

9.1 IA e creazione di valore

Gartner individua tre diverse fonti per la creazione di valore grazie all'applicazione dell'intelligenza artificiale.

- **Esperienza dei clienti** - Le aziende possono migliorare ogni tipologia di interazione con i clienti e facilitarne la crescita e la fidelizzazione. Questo avviene perché l'analisi dei dati con l'IA consente di migliorare la conoscenza dei propri clienti e potenziali clienti, formulare offerte calibrate sul singolo cliente, incrementare i punti di contatto tra organizzazione e cliente e identificare tempestivamente i potenziali problemi (vedere anche il capitolo 10).
- **Nuovi guadagni** - Man mano che l'IA viene utilizzata per migliorare l'esperienza dei clienti, diventa anche più facile vendere prodotti e servizi, consigliando il prodotto o il servizio più giusto nel luogo e nel momento più adatto. L'IA permette anche di identificare nuovi prodotti e servizi da offrire.
- **Riduzione dei costi** - Grazie all'IA si possono ridurre i costi di produzione e di fornitura. La riduzione dei costi è normalmente un modo sicuro per iniziare a sperimentare una qualsiasi tecnologia. Questo perché è più facile prevedere e monitorare una effettiva riduzione dei costi che prevedere e monitorare nuove opportunità di guadagno.

9.2 Classi di applicazioni dell'IA

La maggior parte delle applicazioni dell'intelligenza artificiale finora sviluppate ricadono in una delle seguenti classi (già presentate nel paragrafo 5.1), a cui corrisponde l'utilizzo di specifici strumenti e algoritmi:

- *natural language processing* (NLP), con tecnologie di blog mining (FaceBook, ecc.) per le analisi dei testi;
- *conversational AI* (parte dell'NLP), per la comunicazione con i clienti (customer care, infotainment); essa comprende i chatbot, generalmente basati su testo e programmati per rispondere solo a una determinata serie di domande, e gli assistenti virtuali (*virtual assistant*), capaci di rispondere anche a domande non previste, fino a simulare una conversazione vera e propria;
- *computer vision* o *image processing*, soprattutto negli ambiti medicale, chirurgico, assicurativo (elaborazione pratiche in caso di incidenti automobilistici) e del commercio al dettaglio (distributori automatici);
- *voice recognition*, per riconoscere quanto viene detto da una o più persone;
- *robotic*, ossia robot autonomi (utilizzabili per chirurgia o riparazione delle auto), incluse le auto a guida autonoma;
- *recommendation systems*, utilizzati nell'e-commerce (p.e. Netflix, Spotify e Amazon);
- *intelligent data processing*, con algoritmi che operano su dati (strutturati e non) per estrarre informazioni (sistemi di antifrode, rilevazione delle anomalie, creazione di contenuti, manutenzione preventiva, monitoraggio e controllo, rilevamento di pattern);
- *advanced analytics*, che possono sfruttare algoritmi di ML per eseguire attività di "intelligent data processing.

https://youtu.be/FOMl68X2Amk

"Two Bots Talking: Fake Kirk and A.L.I.C.E."

In questo video troviamo due chatbot: Il falso Kirk contatta il bot A.L.I.C.E. dal 21 ° secolo.

(durata: 3 minuti e 09 secondi)

9.3 Settori dove applicare l'IA

I settori che, fra gli altri, hanno o avranno benefici dell'importante acceleratore dell'IA sono illustrati nei paragrafi successivi:

- marketing e delle vendite;
- utility ed energia;
- bancario;
- finanza;
- assicurazioni;
- telecomunicazioni;
- fornitori di servizi informatici (cloud);
- industria;
- logistica;
- industria automobilistica;
- amministrazione pubblica;
- giustizia e avvocatura;
- sanità;
- militare.

Sono poi presentate alcune applicazioni significative:

- strade intelligenti;
- smart city;
- smart building;
- infrastrutture critiche;
- veicoli autonomi.

Le applicazioni elencate sono esemplificative, ma non esaustive. Ulteriori settori sono quelli della selezione e inserimento del personale in un'organizzazione, della produzione artistica e dello sport.

https://www.youtube.com/watch?v=f_UW1FwJXTY

"Artificial Intelligence in Recruiting"

Questo video spiega come l'utilizzo dell'IA può rendere efficace l'attività di selezione e inserimento del personale.

(durata: 14 minuti e 16 secondi)

https://www.youtube.com/watch?v=19hGiTdA9Vc

"Artificial Intelligence for Human Resources"

Quest video spiega tramite illustrazioni come utilizzare un IA per reclutamento e selezione di risorse.

(durata: 8 minuti e 23 secondi)

https://www.youtube.com/watch?v=rHnOr3oZ5VI

“Visualizing history in 3D.”

Un’IA che rende 3D le foto bidimensionali. Dettagli sula ricerca sono disponibili all’URL: https://shihmengli.github.io/3D-Photo-Inpainting/

Il filmato mostra degli esempi, senza commento audio.

(durata: 1 minuto e 16 secondi)

https://www.youtube.com/watch?v=Ebnd03x137A

“AIVA - "Genesis" Symphonic Fantasy in A minor”

AIVA è un’IA che compone musica, utilizzando un algoritmo basato su architetture di deep learning e reinforcement learning.

Il filmato è un esempio di musica composta da AIVA.

(durata: 2 minuti e 50 secondi)

https://www.youtube.com/watch?v=g8_Qu4gHG5I

“Ecco Ai-da il primo robot umanoide artista ”

Ai.Da, il robot che dipinge.

(durata: 1 minuto e 08 secondi)

https://www.youtube.com/watch?v=j2FwH_pCjhQ

“When AI generated paintings dance to music...”

Il progetto "Neural Synesthesia" che dà vita ai dipinti con la musica.

https://www.musictech.net/news/neural-synesthesia

(durata: 8 minuti e 48 secondi)

https://www.youtube.com/watch?v=nZ950ywJy0M

“Disney’s Stunt Robots Could Change How Hollywood Makes Action Movies”

Utilizzo di robot come stuntman per le scene d'azione dei film Disney.

(durata: 5 minuti e 27 secondi)

https://www.youtube.com/watch?v=byKy9kGnyvo

“Princess Leia Fixed using Deepfakes”

Utilizzo dei Deep-fakes per migliorare il volto in CGI usato per la principessa Leia in “Rogue One”, sostituendolo con il volto vero di Carrie Fisher da giovane.

(durata: 20 minuti e 43 secondi).

https://www.youtube.com/watch?app=desktop&v=h6sA0we-iLc

“AI and Computer Vision based Sports Coaching”

Applicazione di tecniche di visione artificiale e apprendimento per aiutare con l'allenamento sportivo.

(durata: 2 minuti).

Un altro esempio potrebbe riguardare la formazione a distanza, dove l'IA permette di adattare il contenuto di un corso al singolo partecipante anche sulla base delle sue interazioni e, nei sistemi di test da remoto, di individuare infrazioni ai regolamenti (p.e. se sono consultati testi negli esami in cui questo è vietato o se il partecipante riceve suggerimenti da esterni).

10 Settore del marketing e delle vendite

L'IA permette di migliorare le relazioni con i clienti.

L'analisi predittiva, basata su algoritmi IA e serie storiche di dati, già largamente utilizzata da aziende B2C come Amazon, consente di anticipare le esigenze dei propri clienti e potenziali clienti, offrendo loro soluzioni in anticipo o calibrate sui loro specifici interessi, e i potenziali problemi.

Un settore specifico riguarda la presenza online che ogni organizzazione, indipendentemente dal proprio mercato di riferimento o dal suo posizionamento, dovrebbe ritenere necessaria per incrementare i punti di contatto con i propri clienti, comunicare (marketing) e gestire il post vendita.

La presenza online oggi è garantita da un sito web e dalla presenza sui social media. L'IA, come già evidente nel B2C, permette di avere:

- traduzioni pertinenti e multilingua, con contenuti ottimizzati anche per mercati non ancora esplorati;
- dati affidabili e specifici per singolo cliente;
- contenuti originali, aggiornati e ottimizzati per generare traffico, alzando notevolmente il tasso di chiusura delle offerte di vendita.

L'IA permette anche di integrare i CRM con i dati pubblici con cui le persone riempiono il web. Queste funzionalità, per esempio, sono già offerte da Salesforce per i suoi strumenti per la gestione delle vendite.

Il post vendita sta già facendo largamente uso di chat-bot evoluti. Questo permette di ottimizzare i costi, migliorare i tempi di risposta e aumentare le coperture orarie e le lingue gestite.

Ulteriore settore specifico riguarda le chiamate in entrata relative alla fatturazione. In questo caso, sarebbe possibile, per i sistemi di IA, assegnare un punteggio a ogni cliente utilizzando i dati raccolti da diversi sistemi (CRM, contratto, fatturazione, utilizzo, interazioni precedenti con il contact center, il sito web o in negozio, ecc.), prevedere quali clienti sono più propensi a chiamare per le fatture e inviare un video personalizzato con informazioni sui prodotti acquistati, sull'utilizzo dei servizi e sulla fatturazione.

11 Settore delle utility ed energia

Di seguito alcuni esempi di casi d'uso dove l'introduzione dell'IA ha portato dei benefici misurabili e tangibili nel mondo delle utility.

1. Fornire previsioni affidabili per la domanda, il movimento e la fornitura di acqua pulita attraverso l'analisi di dati meteorologici e dati storici sull'utilizzo da parte dei clienti. Sono stati creati scenari "what if" che generano previsioni ogni 30 minuti, sostituendo le precedenti che richiedevano 2-3 settimane di analisi.
2. Determinare la strategia di sostituzione più conveniente per le risorse obsolete degli impianti (vedere il capitolo 23).
3. Visualizzare più facilmente e velocemente i problemi di prestazione attraverso l'analisi complessiva dei dati provenienti dai sensori di gas, elettricità e acqua.
4. Migliorare il servizio ai clienti tramite contact center (vedere capitolo 10).

11.1 Produzione di energie rinnovabili

Nel campo delle utility, il machine learning trova un proficuo utilizzo nelle previsioni a breve termine della produzione di energie rinnovabili, in particolare di tipo solare ed eolico.

Gli effetti meteorologici (nuvolosità e precipitazioni che influenzano l'irradiazione solare nel caso degli impianti solari e velocità e direzione del vento nel caso dell'eolico[55]) possono alterare anche in modo significativo la quantità prodotta e immessa nella rete di distribuzione, a differenza delle fonti tradizionali (p.e. combustibili fossili e nucleare), pianificabile con estrema precisione. I dati meteo storici, ampiamente disponibili, uniti ai dati reali di produzione possono concorrere alla previsione della capacità di produzione.

La Germania, per esempio, sta usando il machine learning per un sistema di early warning che prende dati in tempo reale da impianti eolici e pannelli solari distribuiti sul territorio nazionale per predire l'energia che sarà generata nei due giorni successivi.

Il Governo dell'India, nel report sulla "*National Strategy for Artificial Intelligence*" del 2018[56], individua nella produzione delle energie rinnovabili l'applicazione dell'IA per migliorare l'affidabilità e la convenienza dell'energia fotovoltaica.

Negli Stati Uniti, il National Center for Atmospheric Research (NCAR) ha di recente aggiornato il modello di previsione dell'energia eolica, migliorando l'integrazione tra la previsione

[55] https://pexapark.com/blog/technology/renewable-energy-machine-learning-artificial-intelligence/; https://www.sciencedirect.com/topics/engineering/renewable-energy-forecasting.
[56] https://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf.

numerica del tempo e i metodi di apprendimento automatico. Il nuovo sistema fornisce previsioni a breve termine per l'impegno unitario e il dispacciamento, quantificando l'incertezza della previsione della velocità del vento e la probabilità di eventi estremi come il ghiaccio.

In Italia, l'ente GSE segnala, nel suo report "*Rapporto delle attività 2019*"[57], l'entrata in esercizio nel 2019 del "*primo modello di previsione tramite machine learning applicato all'aggregato, costituito dagli impianti non rilevanti fotovoltaici in regime di ritiro dedicato*".

Terna, nel suo Piano di innovazione, invece, intende utilizzare le tecnologie di advanced analytics e machine learning nell'ambito della sicurezza, stabilità e controllo del sistema elettrico[58].

https://www.youtube.com/watch?v=pghjLyAmc5g

"AI in Renewable Energy: How Is It a Game Changer?"

Video con audio in inglese, che elenca i possibili usi dell'IA nel mercato dell'energia rinnovabile, elencando anche casi di aziende che hanno già messo in pratica queste strategie.

(durata: 3 minuti e 8 secondi)

11.2 SMART grid

Da diversi anni la rete elettrica integrata (*grid*) è un sistema che può essere controllato nel suo insieme attraverso tecnologie digitali.

Recentemente le tecnologie IA sono state introdotte per migliorare il bilanciamento tra le varie componenti della produzione introducendo sistemi che permettono di prevedere, in tempo reale, picchi o cali di consumo in determinate zone o in particolari situazioni. Il sistema di controllo intelligente cerca quindi di attingere dal produttore più vicino e con maggiore disponibilità. Questo permette anche di soddisfare meglio le esigenze degli utenti micro produttori, in grado di rivendere alla rete elettrica la propria energia in eccesso.

57 https://www.gse.it/documenti_site/Documenti%20GSE/Rapporti%20delle%20attivit%C3%A0/RA2019.pdf.

58 https://www.terna.it/it/sistema-elettrico/innovazione-sistema/progetti-innovazione.

11.3 Acquedotti e fognature

Nell'ambito degli acquedotti e delle fognature (*water and waste water*) si stanno sviluppando interessanti tecnologie volte a risolvere problemi che affliggono il servizio idrico come le perdite, la gestione della complessità e la ramificazione del sistema di tubazioni e delle fonti di approvvigionamento, la variabilità degli effetti delle situazioni meteorologiche e molto altro.

Una nuova generazione di sistemi informatici di controllo e di sensori intelligenti già permette di rilevare le perdite occulte e le acque parassite, gestire da remoto, in modalità semi-autonoma o autonoma, gli impianti di acquedotto e fognatura, monitorare in tempo reale gli sfioratori e i sollevamenti fognari oltre che i parametri di qualità dell'acqua erogata.

Sono invece ancora in fase di sviluppo le tecnologie IA per il governo di questi sistemi e sensori, vista la loro eterogeneità, anche se alcuni progetti sono già attivi.

11.4 Mezzi di servizio

Alcune attività svolte dalle società di servizi pubblici richiedono l'impiego quotidiano e capillare sul territorio di vaste flotte di mezzi, basti pensare ai servizi di igiene urbana o agli interventi svolti dai tecnici della rete elettrica. Per questo sono in uso da diverso tempo sistemi, basati su algoritmi di IA, che permettono un'efficiente gestione dei mezzi di servizio attraverso il miglioramento della qualità e dei tempi di risposta del servizio fornito e la riduzione degli eventuali sprechi (consumo di carburante, usura dei mezzi, tempi di trasferimento).

Nella letteratura scientifica, questo genere di problemi viene indicato con l'espressione *vehicle routing problem* (VRP)[59]. Nella sua formulazione più semplice, quella che considera un unico mezzo di capienza infinita, questo problema è noto come *problema del commesso viaggiatore* o *traveling salesman problem* (TSP)[60], molto famoso nell'ambito della logistica. Il mezzo deve transitare per tutti i punti intermedi indicati, prima di far ritorno al punto di partenza, compiendo il tragitto a minima percorrenza. Nonostante la semplice formulazione, all'aumentare del numero di tappe intermedie, la soluzione del problema diventa sempre più complessa in quanto aumenta in modo esponenziale il numero di possibili combinazioni con cui è possibile percorrere l'intero tragitto. Pertanto, anche avendo a disposizione una grande potenza computazionale, spesso non è possibile esplorare tutte le possibili soluzioni. A complicare ancora di più le cose vi è il fatto che spesso è necessario allocare le tappe da visitare a mezzi diversi, ciascuno potenzialmente con diverse caratteristiche (portata, consumi, efficienza, orario di lavoro del personale) e magari non tutti i mezzi sono adeguati allo svolgimento di un

[59] Golden, Bruce L., Subramanian Raghavan, and Edward A. Wasil, eds. The vehicle routing problem: latest advances and new challenges. Vol. 43. Springer Science & Business Media, 2008.
[60] Reinelt, Gerhard. TSPLIB—A traveling salesman problem library. ORSA journal on computing 3.4. 1991, pp. 376-384.

particolare lavoro presso una certa tappa. Gli interventi inoltre possono avere criticità temporali diverse.

Per superare questi problemi, per i sistemi attualmente in uso sono state sviluppate delle tecniche euristiche che sfruttano principi di IA, in grado di limitare in maniera intelligente il numero di soluzioni da verificare.

In base all'approccio utilizzato, si possono distinguere diverse categorie di algoritmi, tipicamente basati su tecniche iterative. Alcuni partono da soluzioni più o meno casuali e cercano iterativamente di migliorarle apportando piccoli cambiamenti e verificandone la bontà[61], altri scompongono il problema in due parti, allocando prima i diversi punti di lavoro ai vari mezzi e procedendo in un secondo momento all'ottimizzazione del percorso di ciascun mezzo[62].

Queste soluzioni richiedono di ricorrere a strumenti per il calcolo delle distanze come quelli offerti da Google[63] o OpenStreetMap[64].

[61] Al-Mulhem, Muhammed, Tareq Al-Maghrabi. Efficient convex-elastic net algorithm to solve the Euclidean traveling salesman problem. *IEEE Transactions on Systems, Man, and Cybernetics*, Part B (Cybernetics) 28.4. 1998, pp. 618-620.
[62] Billy E Gillett e Leland R Miller. A heuristic algorithm for the vehicle dispatch problem. *Operations research* 22.2. 1974, pp. 340–349.
[63] https://developers.google.com/maps/gmp-get-started, last visited 01/10/2020.
[64] Huber, Stephan, and Christoph Rust. "Calculate travel time and distance with OpenStreetMap data using the Open Source Routing Machine (OSRM)." The Stata Journal 16.2 (2016): 416-423.

12 Settore bancario

Le banche dispongono di una grande quantità di dati che oggi, pur essendoci esperienze significative, sono raramente usati per profilare al meglio i propri clienti o per migliorare i processi di governance e compliance.

Nel seguito sono presentati alcuni esempi di applicazione dell'IA nel settore bancario.

12.1 Analisi vendite e acquisti

Un progetto attivo da molti anni, e che quindi si basa su una notevole profondità storica, prevede un'applicazione che memorizza le transazioni dei propri clienti in un data warehouse aziendale.

Per le transazioni dei pagamenti con carte di credito, è stato introdotto e sperimentato un motore di intelligenza artificiale che ha prodotto un modello, continuamente aggiornato, della previsione futura degli acquisti di un cliente (p.e. in quale periodo del mese il cliente spenderà presso il singolo venditore con pagamento elettronico).

L'obiettivo finale (per cui non sono ancora attivate le funzionalità) è un servizio per i venditori che mette loro a disposizione analisi e previsioni per poter meglio gestire la liquidità corrente.

12.2 Piattaforme di pagamento

Ant Group è una società affiliata ad Alibaba. Essa offre la piattaforma di pagamento mobile Alipay e gestisce una straordinaria varietà di attività, tra cui prestiti al consumo, gestione fondi, assicurazioni sanitarie, servizi di rating del credito e persino un gioco online. Ha più di un miliardo di clienti, 10 volte più delle grandi banche USA, con meno di un decimo dei dipendenti. Nel 2018 ha avuto una valutazione di 150 miliardi di dollari, circa la metà di JPMorgan Chase.

A differenza delle banche tradizionali, Ant Group è costruita intorno a un nucleo digitale. Le attività primarie sono interamente gestite dall'IA, anche per approvare prestiti e fornire consulenza finanziaria.

12.3 Segnalazioni frodi con carte

Il sistema di segnalazioni frodi, applicato agli acquisti da Internet, analizza le transazioni fatte con carte debito e credito e le memorizza in un data warehouse. Il progetto è attivo da molti anni e la profondità storica contribuisce all'affidabilità del sistema.

Si è sviluppato un motore basato su regole di supporto all'autorizzazione alle transazioni. Per migliorarlo, sono state acquisite informazioni comportamentali relative all'accesso al dispositivo, compresa l'autenticazione (login) e la sua geolocalizzazione; il motore è continuamente addestrato anche usando i dati delle frodi rilevate.

12.4 Segnalazioni frodi da online banking

La soluzione di segnalazione frodi (*fraud detection*) in tempo reale per le transazioni dei servizi di online banking prevede la profilatura dell'utilizzatore del servizio mediante analisi comportamentale basata sull'operatività dell'utente, i dispositivi usati, il browser, elementi di tipo biometrico (p.e. la velocità di movimento del mouse, la velocità di scrittura, i "movimenti" dello smartphone) e il geo-posizionamento.

Da una base di riferimento (*baseline*), relativa all'intera popolazione o a un singolo utente, è possibile assegnare a ogni richiesta (per esempio login, interrogazione o dispositiva) un grado di rischio e considerarla genuina o sospetta. La banca, in base alle sue regole, può decidere se bloccare o meno un'operazione oppure richiedere ulteriori elementi di riconoscimento all'utilizzatore.

Questi sistemi di DSS (*decision support system*) forniscono una valutazione di rischio in tempo reale.

Alcune limitazioni all'accesso di informazioni, come l'identificativo della SIM Card e l'IMEI del cellulare, attuate dai produttori dei sistemi operativi come Android e iOS, riducono le capacità di contrasto di alcuni schemi di frode come il SIM Swap[65].

12.5 Protezione delle applicazioni

La soluzione, che sfrutta algoritmi di ML, permette di fornire informazioni molto precise sull'attendibilità dell'utilizzatore, la potenziale clonazione del dispositivo, la presenza di malware. Questo permette alla banca di proteggere la propria applicazione da usi non

[65] https://www.certfin.it/educational/minacce-informatiche/sim-swap/.

consoni, ad esempio con funzionalità di runtime application self-protection (RASP), e di proteggere il cliente da possibili frodi.

13 Settore finanziario

Nel trading finanziario sono utilizzate molte tecniche derivate dall'IA[66]. Tra di esse:

- *signal processing*, che opera filtrando i dati per eliminare gli elementi di disturbo e osservare le linee di sviluppo di un mercato;
- *market sentiment*, dove il computer viene lasciato del tutto ignaro delle operazioni in corso fino a che l'algoritmo specifico viene messo all'opera; allora la macchina percepisce subito i comportamenti della domanda e dell'offerta;
- *news reader*, un programma che impara a leggere i principali fenomeni sociali e politici;
- *pattern recognition*, un algoritmo che insegna alla macchina ad imparare e a reagire quando si mostrano, nei mercati, dei tratti che permettono immediati guadagni.

Marble Bar Asset Management (MBAM), società di investimento londinese, ha sviluppato una piattaforma chiamata RAID (Research Analysis & Information Database) per aiutare i gestori di portafoglio a filtrare elevati volumi di informazioni su eventi aziendali, sviluppi di notizie e movimenti di borsa.

Il software identifica gli strumenti finanziari secondo i criteri che soddisfano la propensione di rischio predefinita e ne traccia l'andamento mentre, attraverso l'elaborazione del linguaggio naturale, riconosce e valuta le notizie più significative relative all'investimento per anticipare eventuali fluttuazioni di mercato.

L'Intelligenza artificiale può ottimizzare le performance dei gestori: alcuni sono più avversi alle perdite di altri, trattenendo gli investimenti sottoperformanti più a lungo di quanto dovrebbero, altri potrebbero essere troppo sicuri di sé e prendere eccessivi rischi. L'IA identifica questi comportamenti e suggerisce come migliorare le decisioni.

Infine, l'intelligenza artificiale è usata anche nel mercato immobiliare. Un algoritmo, elaborato da una società tedesca, estrae automaticamente i dati più significativi dalle transazioni immobiliari. A Singapore l'intelligenza artificiale viene utilizzata per calcolare il valore di una proprietà immobiliare, con un mix di algoritmi e di analisi di mercato comparativa.

[66] Giancarlo Elia Valori. "Come l'Intelligenza Artificiale modifica il settore finanziario". Da formiche.net. 2020. https://formiche.net/2020/02/valori-banche-intelligenza-artificiale-finanza/.

14 Settore assicurativo

Lo sviluppo di tecnologie ha portato, anche nel settore assicurativo, a numerosi nuovi strumenti che contribuiscono a garantire efficienza e una maggiore redditività.

Il termine "*InsurTech*" è usato per caratterizzare l'adozione recente delle tecnologie di IA da parte del settore assicurativo, allo scopo di fornire ulteriori possibilità agli assicuratori, fino a concepire una “piattaforma digitale assicurativa” che potrebbe evolversi e operare come Amazon o eBay, incorporando vari prodotti assicurativi e transazioni correlate e, così facendo, favorire la concorrenza su prezzo, servizi e altri fattori utili per una corretta gestione delle pratiche assicurative e la salvaguardia da frodi.

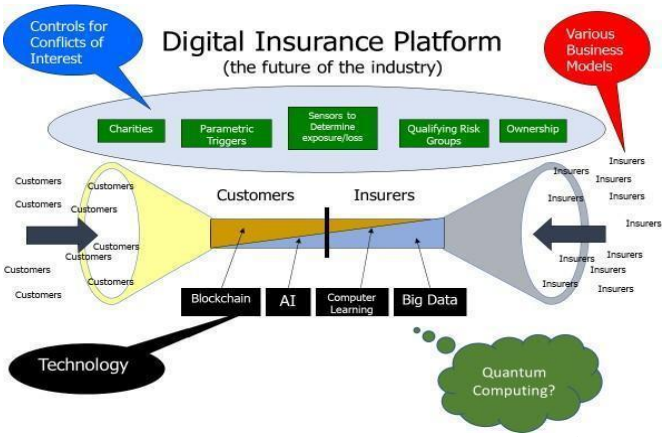


Figura 10 - Piattaforma digitale assicurativa[67]

Più nel dettaglio, l'IA è utilizzata dagli assicuratori (soprattutto all’estero) in diversi modi.

- **Miglioramento del rapporto con clienti -** L’IA, oltre a quanto già specificato nel capitolo 10, permette di costruire rapporti che tengono sempre più conto dell’effettiva condotta dell’assicurato: i comportamenti virtuosi sono premiati con offerte personalizzate, prodotti calibrati e scontati, mentre gli utenti più a rischio o meno performanti sono incentivati a tenere comportamenti più virtuosi e, una volta migliorata la loro condotta e dunque diminuito il loro tasso di rischio, premiati.
- **Automatizzazione della liquidazione dei sinistri auto –** Attualmente i liquidatori dei sinistri sono tenuti a decidere se inviare un'auto in riparazione o liquidare il danno immediatamente in base alle informazioni fornite dal cliente. Con l’utilizzo dell’IA, in particolare con strumenti di data analytics basata su IA e machine learning, il processo di liquidazione dei sinistri e il controllo delle frodi assicurative sono automatizzati sino a fornire la valutazione dei probabili costi di riparazione in tempo reale. L'algoritmo esamina le fotografie del danno presentato dal proprietario; le dimensioni dell'immagine sono standardizzate e vengono convertite in scala di colori di grigio ed in dati binari per consentire all'algoritmo di riconoscere i contorni e, quindi, valutare il livello di danno fino a scegliere l’officina da utilizzare. Se, attualmente, per verificare un sinistro un team di

[67] Figura da Jack E. Nicholson. “Challenges for the Insurance Industry in the Future”. 2019.

persone lavora per settimane per ricavare un dato che ha un'approssimazione di verità del 67-70%, grazie all'IA si stima di arrivare a una precisione intorno al 90% e, così facendo, si può arrivare a espletare l'85% delle liquidazioni danni il giorno stesso della richiesta, passando così da un tempo medio di liquidazione di 28 giorni a pochi secondi. Questi processi automatici sono soggetti al rischio di ricevere input non previsti (paragrafo 33.1.2) perché un malintenzionato potrebbe inviare immagini e video contraffatti (*deepfake*), che la stessa IA potrebbe tentare di individuare mediante analisi antifrode e servizi specifici.

- **Predizione delle frodi assicurative -** L'analisi di eventi precedenti e della sequenza di determinati eventi aiuta a definire i modelli di comportamento e fornisce informazioni sulle possibili azioni fraudolente dei potenziali clienti prima ancora della sottoscrizione di polizze. Analizzando gli eventi è possibile ottenere dati e parametri di ogni interazione storica tra cliente e compagnie assicuratrici, in modo da effettuare un'analisi predittiva e classificare il nuovo cliente secondo i parametri in termini di basso, medio e alto rischio e alto, medio e basso valore.
- **Uso dei dispositivi indossabili -** L'uso di tecnologie indossabili (wearable) – p.e. braccialetti fitness che si basano su IA e applicazioni di IoT - per valutare il benessere e la salute di un cliente, porterà probabilmente un migliore flusso di informazioni sui clienti. Ciò avverrà attraverso il monitoraggio, la raccolta e l'analisi in tempo reale dei dati comportamentali e del rischio. Questo ha anche implicazioni etiche (oltre che legali, considerando la regolamentazione del trattamento di dati personali), in quanto ci si domanda se sia lecito o etico aumentare il premio per l'assicurazione malattia di un individuo perché è stato meno "attivo" per alcuni mesi o si rifiuta di indossare un dispositivo di rilevazione della salute.

15 Settore delle telecomunicazioni

Le aziende del settore delle telecomunicazioni sono state fra le prime a investire risorse per l'applicazione dell'intelligenza artificiale. Sono infatti diverse le applicazioni che negli anni sono state sviluppate e applicate, anche se non sempre con gran successo.

Le soluzioni di IA nel settore delle telecomunicazioni sono finalizzate a diversi compiti e sono state rese note alcune esperienze significative come specificato nei due paragrafi successivi.

15.1 Obiettivi dell'IA nel settore delle telecomunicazioni

L'IA può essere usata per più compiti nel settore delle telecomunicazioni[68].

- **Migliorare la relazione con il cliente,** come già specificato nel capitolo 10.
- **Evolvere la rete (network transformation and automation)**: L'IA automatizza i processi operativi abilitando funzionalità di auto-apprendimento (*self-learning*) e di adattamento dinamico (*adaptive networks*) delle configurazioni delle nuove reti virtuali in funzione dei requisiti dei dispositivi, delle relative applicazioni e delle condizioni operative (p.e. presenza di interferenze radio, caratteristiche del traffico). L'IA permette di affrontare la crescente sofisticazione ed eterogeneità delle tecnologie impiegate (p.e. *software defined networking, network function virtualization, cloud* ed *edge computing*), e l'elevato numero dei dispositivi connessi, ciascuno con specifici requisiti prestazionali.
- **Aumentare il livello di cybersecurity e ridurre le frodi**: l'intelligenza artificiale viene applicata per ridurre le minacce informatiche e identificare i tentativi di frode (capitolo 38).

[68] L. Artusio e altri. "Telcos' application of Artificial Intelligence". Telecom Italia. Presentazione interna del 28 maggio 2019. Disponibile su: https://www.gruppotim.it/content/tiportal/it/notiziariotecnico/edizioni-2019/n-2-2019/N2-Ruolo-Intelligenza-Artificiale-nelle-reti-di-prossima-generazione.html.

15.2 Casi reali

15.2.1 SK Telecom e il monitoraggio delle prestazioni della rete

L'operatore sud-coreano SK Telecom ha introdotto soluzioni di "end-to-end operational intelligence" attraverso l'uso di tre diversi software tra i cui compiti vi sono il monitoraggio della rete radio e della qualità di funzionamento, il rilevamento delle anomalie in tempo reale, l'ottimizzazione e attuazione di misure proattive al fine di anticipare malfunzionamenti e garantire una qualità stabile (calcolata con indicatori di prestazione).

15.2.2 AT&T e ECOMP

AT&T ha conseguito la "hyper-automation" delle proprie reti con tecniche e algoritmi di IA in grado di apprendere (*self-learn*), attuare e modificare dinamicamente e autonomamente (*self-evolve*) le regole di funzionamento della rete. La piattaforma ECOMP (*enhanced control, orchestration, management & policy*) offre funzionalità per l'automazione dei processi di progettazione, creazione e gestione del ciclo di vita dei prodotti e servizi della rete virtuale[69].

Si realizza così la valutazione delle condizioni della rete in tempo reale, delle richieste di traffico e della disponibilità delle risorse per determinare come instradare il traffico per migliorare la qualità del servizio e l'utilizzo delle risorse, in base a regole predefinite.

Le due principali componenti architetturali permettono di:

- abilitare la progettazione e la programmazione delle risorse, dei servizi e dei prodotti supportati dalla rete;
- eseguire le logiche definite in fase di progettazione, con una componente DCAE (*data collection, analytics and events*) che offre le funzionalità di data analytics (raccoglie dati prestazionali, d'uso e configurazione degli elementi della rete) mentre il modulo A&AI (*active and available inventory*) fornisce una vista in tempo reale delle risorse, servizi e prodotti e delle loro relazioni.

[69] https://www.gruppotim.it/content/tiportal/it/notiziariotecnico/edizioni-2019/n-2-2019/N2-Ruolo-Intelligenza-Artificiale-nelle-reti-di-prossima-generazione.html.

15.2.3 NTT e i CAT

NTT ha sviluppato un sistema di raccolta dati, diagnosi e controllo automatico dello stato della rete basato su cicli procedurali CAT (*collect-analyze-test*)[70].

L'architettura prevede dei motori distribuiti ("*policy-driven CAT cycles*") preposti alla raccolta e analisi dei dati generati da ciascun elemento di rete (fisico e logico). La conoscenza acquisita, e continuamente aggiornata, si basa sulla corrispondenza tra stati di funzionamento e pattern di dati. In particolare, le variazioni nel tempo dei dati raccolti e lo stato di un determinato elemento di rete vengono continuamente confrontate con dei modelli di riferimento, evidenziando scostamenti e anomalie.

[70] A. Hirano, "Autonomous network diagnosis with AI", da "Proceedings of Optical Fiber Communications Conference". Tu2E.4. 2019.

16 Fornitori di servizi informatici (cloud)

Intelligenza artificiale e cloud computing possono essere usati insieme perché da una parte l'IA può rendere il cloud più efficiente e dall'altra il cloud permette di supportare le soluzioni di IA.

16.1 Il cloud usa l'IA

Il "cloud intelligente" è un'infrastruttura cloud in grado di prevedere le tendenze (utilizzo delle risorse, accesso ai dati, avvio delle applicazioni e utilizzo delle informazioni, comportamento degli utenti, ecc.). e capace di decidere autonomamente come distribuire le risorse. Al momento, però è ancora in fase embrionale.

Di contro, l'intelligenza artificiale, già oggi, attraverso l'*intelligent data processing*, può fornire utili risposte e analisi dei dati che potrebbero rendere il cloud computing non solo efficiente ma anche efficace rispetto a particolari esigenze. In questo caso si tratta di tecnologie non ancora utilizzate pienamente o ancora in fase di ideazione.

L'esigenza più importante riguarda i trattamenti di grandi moli di dati, come quelli che interessano l'e-commerce, il trading finanziario e il risk management e come quelli raccolti dai sempre più numerosi dispositivi che costituiscono l'internet of things (IoT) e dalle app per dispositivi mobili. Alcune di queste applicazioni sono oggetto degli altri capitoli relativi all'uso dell'IA.

La sicurezza dei sistemi informatici, come meglio specificato nel capitolo 42, può giovarsi dell'intelligenza artificiale. In particolare, per gestire le informazioni rilevate dai sistemi di sicurezza e gestire la sicurezza dei servizi cloud di diversi vendor, già da diversi anni sono a disposizione piattaforme CASB (cloud access security brokers). Alcuni CASB utilizzano algoritmi di machine learning per:

- proteggere gli accessi ai servizi, quindi le autenticazioni, come nel caso degli spray password attack;
- contrastare tentativi di esfiltrazione di dati da parte di attaccanti esterni o di dipendenti infedeli;
- identificare ATP.

16.2 L'IA usa il cloud

Molti fornitori di servizi cloud (CSP) stanno introducendo istanze di calcolo specializzate che permettono di gestire gli elevati carichi di lavoro necessari per integrare l'intelligenza artificiale nelle applicazioni.

I pesi massimi del cloud, ma non solo loro, hanno lanciato servizi intelligenti: Amazon integra Alexa alle applicazioni aziendali di Amazon Web Services e ha lanciato diversi servizi di machine learning e Google e Cisco hanno ampliato la loro offerta di intelligenza artificiale, puntando a implementazioni cloud ibride.

Microsoft sta facendo passi per aiutare gli sviluppatori di IA a costruire applicazioni da eseguire sul proprio cloud Azure. Il progetto Brainwave consente agli sviluppatori di utilizzare i cosiddetti *gate-programmable field array* (FPGA), che consentono ai modelli IA di essere più veloci e riconfigurabili anche dopo l'installazione su un server Azure.

IBM ha reso IBM Watson utilizzabile su qualsiasi cloud, permettendo alle imprese di superare il limite del vendor lock-in e iniziare a usare soluzioni di intelligenza artificiale.

Il cloud può anche supportare efficacemente il meccanismo di addestramento dell'intelligenza artificiale, grazie all'elevata mole di dati che può mettere a disposizione. Il cloud permette quindi di fornire soluzioni di "intelligenza artificiale as-a-service".

Per esempio, IBM promuove Storage Insights, piattaforma cloud con funzioni di IA, per la gestione degli storage. I dati sull'utilizzo e sulle prestazioni degli storage IBM sono caricati sul cloud di IBM, dove sono analizzati utilizzando algoritmi di apprendimento automatico per determinare le impostazioni ottimali a seconda degli obiettivi di prestazione e di prezzo del cliente.

17 Settore industriale

L'IA può contribuire in modo significativo a una trasformazione radicale dell'industria manifatturiera nei prossimi anni, integrando e aumentando l'efficacia delle altre tecnologie che caratterizzano il paradigma dell'Industria 4.0 (industrial IoT, big data and advanced analytics, cloud computing, robotica collaborativa, realtà aumentata e virtuale, stampa 3D, simulazione e sistemi cyber-fisici, integrazione digitale dei processi, cyber security, eccetera).

In base ai risultati di una recente ricerca[71], che ha analizzato i livelli di adozione dell'IA da parte di 300 aziende manifatturiere operanti a livello globale, individuando 22 casi d'uso diversi, raggruppati in 7 aree funzionali (dallo sviluppo prodotti, alla produzione, alla manutenzione), l'Europa è in una posizione di leadership, con oltre la metà delle aziende rappresentative che ha adottato almeno uno dei casi d'uso.

Tra i diversi possibili casi d'uso, i seguenti tre risultano essere i più indicati come punto di partenza per l'avvio di un percorso di adozione dell'IA in ambito manifatturiero:

- manutenzione predittiva;
- controllo della qualità dei prodotti;
- pianificazione della domanda.

A ciò contribuiscono alcune caratteristiche:

- chiarezza dei benefici;
- relativa facilità di implementazione;
- potenziale disponibilità di dati (prestazioni di macchinari ed apparati, immagini e video dei prodotti, eccetera);
- disponibilità di adeguate conoscenze sulle soluzioni esistenti;
- possibilità di aggiungere funzionalità che aumentano la visibilità e la comprensibilità degli algoritmi di IA, consentendo al personale di capire meglio come vengono prese le decisioni e rendendo più semplice l'adozione della soluzione.

https://www.youtube.com/watch?v=TeIKS5MEluU

"Reaching Industry 4.0 with AI"

Video riepilogativo, senza audio ma con testo in inglese in sovraimpressione, delle potenzialità dell'IA nel futuro dell'industria 4.0.

(durata: 1 minuto e 18 secondi)

[71] Capgemini. *Scaling AI in Manufacturing Operations: A Practitioners' Perspective*. 2019

Altri casi d'uso dell'intelligenza artificiale sono orientati al miglioramento produttivo attraverso elementi oggetto dei successivi paragrafi.

17.1 Personalizzazione delle produzioni

Elementi di IA permettono, con tecnologie oggi utilizzate in più contesti, di personalizzare maggiormente i prodotti nel tentativo di avvicinarsi il più possibile alle esigenze dei singoli clienti, in una prospettiva di prodotto quasi artigianale. Questo permette di incrementare le vendite e fidelizzare la clientela. L'IA permette di gestire il grande aumento di lotti produttivi e la conseguente riduzione delle parti per lotto.

17.2 Ottimizzazione dei processi produttivi

Sono sempre di più le linee produttive o le macchine con sistemi intelligenti in grado di migliorare autonomamente l'efficienza dei processi di produzione. I dati raccolti provengono dal monitoraggio delle quantità di materie prime, dei tempi del ciclo produttivo, delle temperature, dei tempi di realizzazione, degli errori e scarti e dei tempi di fermo macchina. Attraverso l'analisi delle decisioni degli operatori, l'IA impara a prendere decisioni autonome o supportare gli operatori.

Esempio significativo sono i *co-bot* o *collaborative robot,* impiegati per automatizzare, rendendole più efficienti, le attività produttive troppo pesanti o pericolose per i lavoratori umani. Rispetto al passato, questi robot hanno un elevato livello di interazione con l'operatore umano, lavorando al suo fianco durante le varie fasi della filiera, sorvegliando il corretto funzionamento degli apparati, eseguendo test e imballando i prodotti finiti. Alcuni di questi robot sono veicoli a guida autonoma (vedere anche il paragrafo 24.2.1).

Per il controllo qualità sono disponibili sistemi di intelligenza artificiale *computer vision,* sistemi di analisi visiva utilizzati per una accurata verifica della presenza di difetti.

17.3 Approvvigionamento automatico di materiale

Dal monitoraggio dei bisogni nelle varie fasi di lavorazione, condizionati da eventuali anomalie o malfunzionamenti, i sistemi IA oggi in uso sono in grado di emettere richieste di approvvigionamento autonomamente in base alle reali necessità. In alcuni casi i sistemi di IA possono anche gestire la fase iniziale di quotazione.

17.4 Manutenzione predittiva

La manutenzione continua dei macchinari e delle attrezzature produttive rappresenta una componente di spesa importante nell'industria manifatturiera, con un impatto cruciale sulla profittabilità delle linee di produzione.

La manutenzione predittiva utilizza algoritmi di IA (machine learning e reti neurali), insieme a digital twins e analitiche, per tracciare il profilo del normale comportamento dei sistemi della filiera produttiva e prevedere futuri guasti (*machines failure prediction*). Ciò consente di ridurre i tempi di fermo non pianificati e di estendere la vita utile residua delle macchine e delle attrezzature di produzione.

17.5 Fattori critici

I principali fattori critici di successo per l'adozione su larga scala di tali casi d'uso nei vari processi operativi sono:

- utilizzo dei dati raccolti in tempo reale dall'ambiente di produzione, grazie a integrazioni con i sistemi IT (MES e ERP) e con l'infrastruttura IIoT;
- uso di una piattaforma centrale per la memorizzazione e l'analisi dei dati utilizzati dalle applicazioni IA;
- sviluppo di competenze interne in ambito IA;
- monitoraggio continuo delle prestazioni, della qualità dei risultati e dell'affidabilità.

18 Settore della logistica

Negli ultimi anni la rivoluzione digitale ha influenzato molti aspetti dei settori industriali e la cosiddetta Logistica 4.0 è uno dei settori in cui l'utilizzo dell'IA è sempre più diffuso: innovazioni come strade intelligenti e veicoli autonomi sono alcuni dei casi d'uso più promettenti.

Lo scopo principale di molte applicazioni di IA nel settore della logistica è l'automazione delle azioni *time-consuming* e la conseguente riduzione dei costi. Molte aziende tecnologiche hanno investito ampiamente in questa tecnologia, come dimostrano i magazzini automatizzati di Amazon, DHL o Waymo, così come il Porto di Singapore.

Figura 11 - Intelligenza artificiale e sistema di trasporto avanzato[72]

18.1 Magazzini automatizzati

L'IA nell'automazione del magazzino viene utilizzata per prevedere la domanda di determinati prodotti e, quindi, avere una pianificazione con largo anticipo con la conseguente riduzione dei costi di trasporto.

I sistemi di automazione del magazzino e la robotica nella gestione delle merci basati su tecnologie di IA e attualmente disponibili offrono la possibilità di organizzare (e rendere più veloci e precise) molte attività ripetitive, quali il controllo e l'organizzazione dell'inventario e il trasporto tra i magazzini, nonché il controllo di qualità in modo automatico.

[72] Fonte: https://www.shutterstock.com/.

L'IA è anche usata per migliorare l'allocazione delle risorse (p.e.: piani di lavoro, rotazione del personale, rifornimento di punti vendita, riordino di prodotti).

18.2 Veicoli a guida autonoma

L'uso di veicoli automatizzati nel settore della logistica viene visto come una possibilità di risparmiare tempo e denaro e di ridurre gli incidenti. I veicoli senza conducente contribuiscono ad accelerare il processo di consegna, possono ottimizzare i percorsi, ridurre gli incidenti di errore umano e lavorare 24 ore su 24, 7 giorni su 7.

Ulteriori dettagli sui veicoli a guida autonoma sono nel paragrafo 19.2 e, per quelli aerei, nel paragrafo 28.2.3.

18.3 Consegna dell'ultimo miglio

L'obiettivo della consegna del cosiddetto "ultimo miglio" è quello di garantire che le merci raggiungano il cliente finale il più velocemente possibile ("*I want it when I want it*" è l'espressione coniata da DHL in un recente studio)[73].

L'elevato grado di automazione possibile grazie alle tecnologie di IA consente di assecondare questa esigenza in tutte le fasi del processo di consegna, garantendo che il prodotto venga inviato attraverso i canali giusti, confezionato correttamente e che il tempo di consegna sia notificato al cliente.

Droni dotati di IA sono stati impiegati in alcune parti del mondo per la consegna dell'ultimo miglio, aumentando significativamente le possibilità di scelta del cliente circa luoghi e orari di consegna e consentendo comunque al distributore una ottimizzazione dei costi[74] .

Sono in fase di sperimentazione reti di avio trasporto costituite da UAV (*unmanned air vehicle*). Esse potrebbero essere strutturate come segue: le spedizioni che arrivano dall'esterno alle estreme periferie della città vengono ordinate presso le strutture esistenti (hub, magazzini, siti di cross-dock) e le spedizioni che soddisfano determinati criteri vengono separate automaticamente. Oltre a dimensioni, peso e criticità temporale, i criteri decisionali di trasporto potrebbero includere anche metriche dinamiche quali: condizioni stradali attuali, inquinamento atmosferico, carico di rete, condizioni atmosferiche. Ogni UAV preleverebbe automaticamente le spedizioni assegnate da un nastro trasportatore e decollerebbe. Sulla via del ritorno all'hub, l'UAV potrebbe effettuare consegne punto a punto che si trovano sul suo percorso.

[73] Shortening the last mile: winning logistics strategies in the race to the urban consumer. https://www.dpdhl.com/content/dam/dpdhl/en/media-center/media-relations/documents/2018/dhl-whitepaper-shortening-the-last-mile.pdf.
[74] S. Perera, M. Dawande G. Janakiraman V. Mookerjee, "Retail Deliveries by Drones: How Will Logistics Networks Change?", Wiley, 2020.

Nella successiva figura sono mostrati alcuni mezzi UAV utili per tali scopi[75].

	Advantage	Disadvantage	Visual
Fixed-Wing	• Long range • Endurance	• Horizontal take-off, requiring substantial space (or support, e.g., catapult) • Inferior maneuverability compared to VTOL (Vertical Take-Off and Landing)	 Source: Indra Company
Tilt-Wing	• Combination of fixed-wing and VTOL advantages	• Technologically complex • Expensive	 Source: sUAS News
Unmanned Helicopter	• VTOL • Maneuverability • High payloads possible	• Expensive • Comparably high maintenance requirements	 Source: Swiss UAV
Multicopter	• Inexpensive • Easy to launch • Low weight	• Limited payloads • Susceptible to wind due to low weight	 Source: Microdrones

Figura 12 - Tipologia di UAV per il trasporto logistico – corrieri

Le decisioni di instradamento sarebbero sempre dinamiche, il che significa che una rete intelligente ridistribuirebbe tutte le risorse in tempo reale, a seconda del carico e dell'urgenza di determinate spedizioni e di ogni altra variabile di sistema ritenuta notevole per l'ottimizzazione e il bilanciamento della rete logistica di trasporto.

Il caso del primo e dell'ultimo miglio pone preoccupazioni per la privacy e la sicurezza in un ambiente urbano densamente popolato. Ed è il più impegnativo per l'integrazione nelle infrastrutture urbane esistenti.

Il potenziale della tecnologia UAV è evidente ove siano presenti infrastrutture di basso livello o condizioni geografiche difficili (p.e. ambienti montuosi e isole). Per il settore della logistica, la consegna rurale tramite UAV è attraente non solo nelle applicazioni di emergenza ma anche perché le località remote a basso volume rappresentano un elemento di costo significativo della rete di distribuzione.

75 https://www.dhl.com/content/dam/downloads/g0/about_us/logistics_insights/DHL_TrendReport_UAV.pdf.

Per le località remote di un'isola, un possibile “use case” è la consegna di pacchi alle isole vicine alla costa, sostituendo un processo esistente assai complesso che coinvolge auto, barche e lavoratori postali e fornendo al contempo nuovi servizi aggiuntivi.

https://www.youtube.com/watch?v=vL6_P-iWAAM

“3 Applications of Artificial Intelligence in Supply Chain Management”

Video muto che presenta tre esempi di IA applicabile al settore dei trasporti e logistica.

(durata: 50 secondi)

https://www.youtube.com/watch?v=kp0Q6OQUdHM

“Machine learning in supply chain”

Video introduttivo che spiega il potenziale utilizzo del ML nel settore dei trasporti e della logistica.

(durata: 1 minuto e 53 secondi)

19 Settore dell'industria automobilistica (automotive)

Uno dei più conosciuti e discussi sistemi basati su IA nel settore dell'automotive è senz'altro il sistema di assistenza alla guida con funzioni quali guida automatica, frenata automatica, prevenzione delle collisioni, segnalazione di pedoni e ciclisti, avvisi di traffico incrociato e controlli di crociera intelligenti che rendono l'esperienza di guida più sicura.

Nei paragrafi successivi sono descritte applicazioni significative dell'intelligenza artificiale nel settore automotive. L'ultimo paragrafo tratta dei rischi specifici.

19.1 Fabbricazione

Nell'ambito della fabbricazione, le applicazioni sono quelle già viste per il settore industriale (capitolo 17).

19.2 Veicoli a guida autonoma

L'obiettivo dei sistemi di intelligenza artificiale applicata alla guida autonoma è di assicurare un minor numero di incidenti, meno traffico e meno inquinamento.

Nei cosiddetti "automated vehicles" alcune funzioni critiche per la sicurezza del conducente e dei passeggeri, come la sterzata, l'accelerazione o la frenata, avvengono senza un diretto intervento umano. Possono essere:

- autonomi, nel caso usino solo sensori montati a bordo;
- connessi, in questo caso vengono definiti CAV ossia *connected automated vehicles*.

L'essere dotati di connettività consente ai CAV di sfruttare a pieno le funzionalità messe a disposizione dall'intelligenza artificiale e dalle applicazioni in cloud.

Di seguito sono riportati i livelli di automatizzazione secondo la classificazione SAE (Society of automotive engineers).

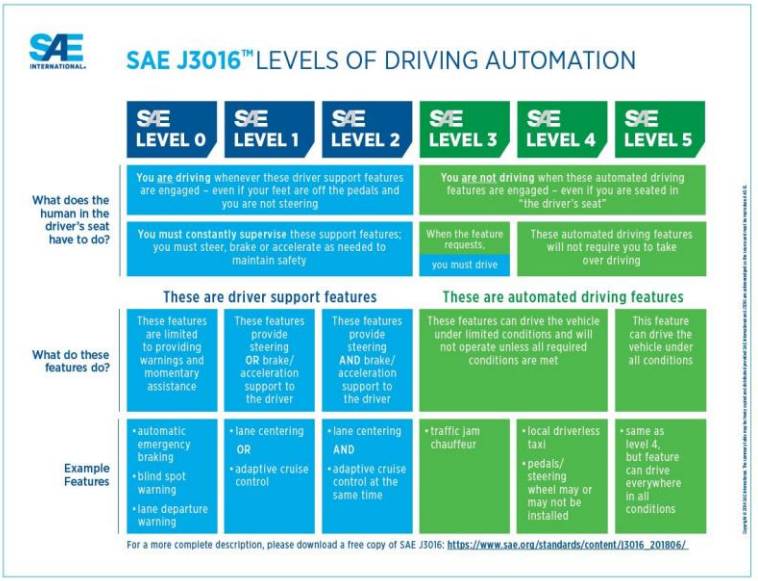


Figura 13 - Livelli SAE[76]

Relativamente ai primi livelli, sono ormai comunemente utilizzati sistemi di supporto alla guida come l'adaptive cruise control (ACC), il park assist, il lane centering assist, i meccanismi di rilevamento della presenza del guidatore e del suo livello di attenzione, l'assistenza alla guida in autostrada e il parcheggio custodito automatizzato (valet parking).

Dal SAE level 3, il sistema prende decisioni autonome e mette in atto azioni sulla base delle informazioni che provengono dal mezzo. Le decisioni possono essere di tipo strategico (ad esempio pianificazione e ottimizzazione dei percorsi), tattico (ad esempio su come districarsi nel traffico, se superare un altro veicolo, se effettuare un cambio di corsia e se cambiare velocità) e operativo (decisioni da prendere "all'ultimo secondo" in merito a correzioni di posizionamento nella corsia, aggiramento di un ostacolo presentatosi improvvisamente, ecc.).

In un veicolo connesso, l'intelligenza artificiale consente un'interazione tra il mezzo, le persone e l'ambiente. Ciò si realizza in quattro momenti.

- *Sense*: l'intelligenza artificiale consente al veicolo di percepire il mondo circostante acquisendo ed elaborando una grande quantità di informazioni come immagini, suoni, parole e testi. Il *machine learning* può elaborare grandi quantità di dati visivi, come immagini di segnali stradali o di strade in diverse condizioni meteorologiche, e può riconoscere un'ampia gamma di oggetti in 2D, comprendere da un punto di vista spaziale scene in 3D, valutare la collocazione e la distanza di un ostacolo, ecc.
- *Comprehend*: l'intelligenza artificiale consente al veicolo di comprendere informazioni raccolte applicando tecnologie di *analytics* e ricavandone il significato.
- *Decide*: tramite un'attività di confronto e classificazione delle informazioni provenienti dai sensori, l'intelligenza artificiale sviluppa un processo di decisionale basato su

[76] https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic#:~:text=The%20J3016%20standard%20defines%20six,5%20(full%20vehicle%20autonomy).&text=Ranch%20in%20Florida.The%20latest%20J3016%20graphic%20is%20a%20living%20document.,J3016%20itself%20evolves%2C%20Pokrzywa%20explained.

machine learning e *deep learning*, ad esempio programmando percorsi, traiettorie e manovre nel traffico.

- *Act*: sulla base dell'analisi effettuata, l'intelligenza artificiale consente al veicolo di compiere delle azioni, sia nel mondo fisico che in quello digitale.
- *Learn*: l'intelligenza artificiale consente al veicolo di correggere e migliorare continuamente i propri comportamenti sulla base delle informazioni apprese, dell'esperienza e dei risultati ottenuti precedentemente.

Il *machine learning* è in particolare il fattore critico nello sviluppo di veicoli CAV perché consente al veicolo di comprendere il mondo intorno a sé e gli consente di agire senza essere esplicitamente programmato, ma traendo insegnamento dagli input provenienti dall'esterno e dall'esperienza pregressa.

C'è ancora molto lavoro da fare per raggiungere la piena autonomia dei veicoli e ad ora la tecnologia e le normative non consentono ai veicoli completamente autonomi di guidare su strade senza la supervisione umana. Sono in corso alcune sperimentazioni, tra cui il programma Next[77], sviluppato da una startup in Veneto: un sistema di trasporto pubblico basato su veicoli elettrici a guida autonoma e con moduli separabili che possono ospitare fino a dieci persone.

https://www.youtube.com/watch?v=Fo6pWIi-Ixo

"BMW Factory - Integration of A.I. in the Production Line"

BMW usa l'IA e il riconoscimento visivo per il controllo qualità lungo il percorso della sua catena di montaggio.

(durata: 2 minuti e 56 secondi)

https://www.youtube.com/watch?v=kJD5R_yQ9aw

"Unedited 40 Minute Ride in Mobileye's Autonomous Car"

Un viaggio attraverso le strade di Gerusalemme con un'auto a guida autonoma di Mobileye. Vengono mostrate contemporaneamente la vista interna, la vista dall'alto con un drone, e la visione elettronica dell'IA.

(durata: 25 minuti e 56 secondi)

[77] https://www.corriere.it/tecnologia/automotive/notizie/next-future-transportation-l-autobus-componibile-autonomo-inventato-un-padovano-ff40f2da-457d-11e7-a851-a338795668a8.shtml

19.3 Previsione dei guasti dei veicoli

Diverse aziende utilizzano modelli predittivi per prevedere i guasti delle macchine e metterle in manutenzione prima che si rompano: è possibile monitorare, valutare e segnalare in maniera costante le condizioni di usura dei vari componenti di un veicolo.

Nelle corse sportive un sistema del genere permette di avere vantaggio in pre-gara o in gara perché permette di definire al meglio la strategia di sostituzione dei componenti (p.e. cambio motore e gomme).

19.4 Rilevazione degli incidenti

I veicoli possono montare un sistema di soccorso digitale che rileva automaticamente gli incidenti in auto e lancia un allarme indicando anche la posizione GPS del veicolo. Chi riceve la segnalazione può contattare l'utente in auto per verificare la sua incolumità e attivare i soccorsi.

Per quanto riguarda la rilevazione degli incidenti nelle cosiddette strade intelligenti (*smart road*), vedere il capitolo 24.

19.5 Monitoraggio dello stile di guida del conducente

L'intelligenza artificiale permette di analizzare lo stile di guida del conducente allo scopo di inviare allarmi in caso di eventi pericolosi come deviazione dalla corsia di marcia, collisione imminente e condizioni metereologiche rischiose. Tali sistemi sono anche in grado di esprimere una valutazione della guida in modo da aiutare il conducente a migliorare il proprio comportamento al volante.

19.6 Gestione delle flotte

Sulle flotte, l'impiego dell'intelligenza artificiale permette di tracciare il percorso dei veicoli, ricevere report sulle loro performance e raccogliere informazioni sulla dinamica di eventuali incidenti allo scopo di garantire una maggiore sicurezza dei conducenti e di ridurre i premi assicurativi.

19.7 Servizi di car sharing

Nel *car sharing* si fa utilizzo di IA per rendere le offerte di servizi personalizzate per gli utenti finali. Le aziende sono in grado di personalizzare le tariffe per un determinato utente e offrirgli consigli personali in base al meteo, all'orario e alla destinazione, per cui possono variare la tipologia di auto (city car di giorno, familiare alla sera), marca dell'auto preferita, destinazione (casa, lavoro, ecc.) e anche preferenze musicali.

Grazie a IA e data analytics, le aziende di car sharing possono valutare in quali posti e orari distribuire le auto per aree geografiche.

19.8 Intelligenza artificiale nel settore automotive: i rischi

Il crescente ricorso alle tecnologie di intelligenza artificiale in settori particolarmente critici come l'automotive pone ovviamente in primo piano temi di sicurezza, di conformità normativa e di natura etica. Rimandiamo, per molti di questi aspetti, alla lettura di specifici capitoli di questa pubblicazione.

Sebbene alcune statistiche come quelle del World Economic Forum già qualche anno fa prevedesse una riduzione del 9% degli incidenti stradali entro il 2025 e la possibilità di salvare 900.000 vite grazie all'adozione delle tecnologie di guida autonoma[78], restano tuttavia alcuni aspetti chiave su cui ci si dovrà confrontare nei prossimi anni. Nel paragrafo 38.1.4.1 sono riportati alcuni casi di incidenti che hanno coinvolto automobili a guida autonoma.

Il rischio relativo alla trasparenza dell'IA (capitolo 31) è importante anche per l'automotive perché, nel caso in cui si verifichi un incidente chiaramente attribuibile a una decisione errata del veicolo a guida autonoma, è importante comprendere quale sia stato il processo logico che ha portato a quelle conseguenze. Questo per avviare un intervento tecnico che prevenga l'occorrenza di incidenti simili e un intervento giuridico per l'identificazione delle responsabilità[79].

[78] http://reports.weforum.org/digital-transformation/wp-content/blogs.dir/94/mp/files/pages/files/wef-dti-automotivewhitepaper-final-january-2016-200116a.pdf.
[79] https://www.scientificamerican.com/article/demystifying-the-black-box-that-is-ai/.

20 Settore dell'amministrazione pubblica

Nel corso degli ultimi anni, alcune pubbliche amministrazioni si sono dotate di software di IA. La base giuridica di tale utilizzo può rinvenirsi nell'art. 3-bis della legge 241/1990, nel D. Lgs. 82/2005 (Codice dell'amministrazione digitale) e nei principi di efficienza ed economicità dell'azione amministrativa stabiliti dall'art. 97 della Costituzione.

Il ricorso all'intelligenza artificiale ha indubbi vantaggi per le amministrazioni pubbliche nella riduzione della tempistica procedimentale per operazioni meramente ripetitive e prive di discrezionalità, per l'esclusione di interferenze dovute a negligenza o dolo del funzionario (essere umano) e la conseguente maggiore garanzia di imparzialità della decisione automatizzata.

L'uso dell'IA, però, deve soddisfare alcune caratteristiche evidenziate da alcune normative e nel corso dei provvedimenti giudiziari, come richiamato nei capitoli 30 e successivi.

Il Joint Research Centre (JRC) presso la Commissione europea ha recentemente pubblicato uno studio[80] nel contesto di AI Watch[81]. Esso rappresenta la prima analisi dell'utilizzo della IA nelle pubbliche amministrazioni europee (230 casi applicativi in UE, Norvegia, Svizzera e Gran Bretagna).

Lo studio rileva due tipologie ricorrenti di uso della IA: i chatbot (per la simulazione di conversazioni) e le previsioni basate su dati socioeconomici. Sottolinea inoltre l'interesse sempre più crescente per l'uso della IA per supportare la ridefinizione dei servizi e delle politiche pubbliche e per migliorare la qualità del coinvolgimento dei cittadini.

Lo studio pone particolare attenzione all'impatto etico e sociale e alla necessità di porre il tema dell'IA all'interno degli sforzi di regolazione per bilanciare le elevate aspettative con i potenziali effetti negativi. Lo studio propone inoltre uno schema per la valutazione d'impatto della IA e del suo sviluppo nei servizi pubblici, cercando di definire i fattori cruciali per valutare le diverse applicazioni.

[80] https://op.europa.eu/en/publication-detail/-/publication/4c72dd88-bcda-11ea-811c-01aa75ed71a1/language-en.
[81] https://ec.europa.eu/knowledge4policy/ai-watch_en.

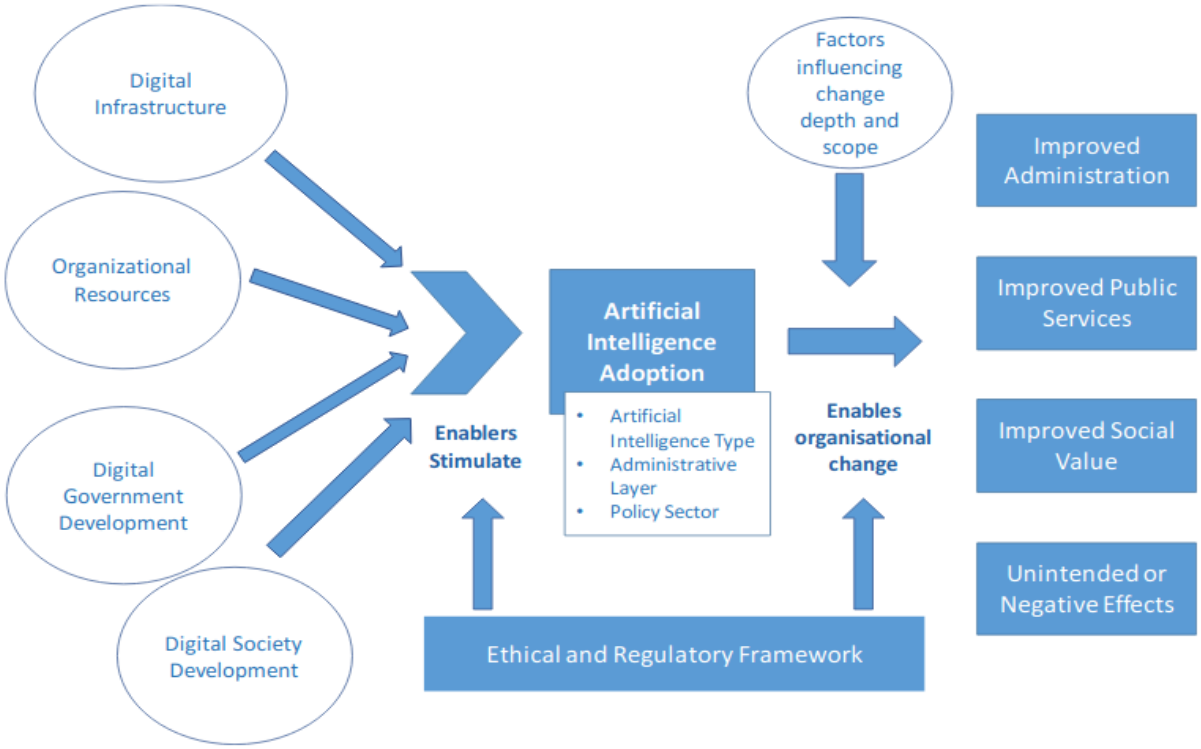


Figura 14 – Schema degli impatti dell'IA nei servizi pubblici[82]

Il report conclude richiamando i principi individuati, con riguardo al settore pubblico, dagli esperti della UE[83]:

1) fornire servizi basati sulla IA e incentrati sull'uomo;
2) considerare il Governo come piattaforma, catalizzando lo sviluppo dell'IA in Europa;
3) fare un uso strategico degli appalti pubblici per finanziare l'innovazione e garantire un'IA affidabile;
4) salvaguardare i diritti fondamentali nei servizi pubblici basati sulla IA e proteggere le infrastrutture sociali.

[82] AI Watch "Artificial Intelligence in public services - Overview of the use and impact of AI in public services in the EU ".
[83] https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence.

21 Giustizia e avvocatura

Studi di psicologia, economia comportamentale e scienze cognitive mostrano che gli esseri umani hanno capacità di ragionamento limitate e imperfette, specialmente quando si tratta di statistica e probabilità.

L'intelligenza artificiale, in questo ambito usata soprattutto negli USA, permette di avvisare il giudice quando la decisione è incoerente con analoghe valutazioni pregresse, quando si discosta da parametri predefiniti o è difforme da una media ponderata delle decisioni prese dai colleghi in situazioni analoghe.

Attraverso il *modelling*, gli avvocati (soprattutto negli USA) possono usare l'intelligenza artificiale per analizzare le decisioni passate del giudice assegnato, simulare l'andamento di un processo e cercare di predirne l'esito.

22 Settore sanitario

Le aree di applicabilità dell'IA in sanità sono oggi molto ampie, dal software con finalità direttamente terapeutiche o diagnostiche, ai software di aiuto al medico nella sua attività professionale fino ad arrivare ai sistemi di gestione informativa ospedaliera, regionale o nazionale.

Molti dei software di intelligenza artificiale sono poi oggi qualificati come dispositivi medici e come tali sottoposti a specifica legislazione (Direttiva 93/42/CEE, recepita in Italia dal D. Lgs. 46 del 1997, e dal Regolamento UE 2017/745).

Le aree di applicabilità dell'IA in sanità possono essere classificate in:

- dispositivi medici, dove maggiori sono le esperienze di applicazione;
- sistemi informatici ospedalieri e regionali.

Ognuna delle possibili applicazioni è caratterizzata da specifici benefici ma anche da potenziali rischi; inoltre è sottoposta a specifici vincoli regolamentari applicabili ai dispositivi medici.

22.1 Dispositivi medici

Per essere considerato dispositivo medico, un apparecchio o strumento deve apportare un beneficio clinico, pur non utilizzando principi farmacologici[84].

[84] Regolamenti 2017/745 e 746 del Parlamento Europeo e della Commissione Europea

Per semplificare l'inquadramento dell'IA nei dispositivi medici, il campo di applicabilità può essere diviso in due categorie.

- I **software standalone** sono indipendenti da un dispositivo medico fisico. Esempi sono i software installabili su un comune PC. Questi software usano l'IA per diversi scopi, tra cui: controllare altri dispositivi medici, ottimizzandone i parametri di funzionamento sulla base dei dati storici e di quelli specifici del paziente; fornire assistenza nel percorso diagnostico e terapeutico del paziente; fornire indicazioni cliniche analizzando dati fisiologici; proporre al medico un'opinione aggiuntiva.
- I **software embedded** sono installati sui dispositivi medici e sono dedicati al controllo dei dispositivi stessi, ottimizzandone parametri e funzionamento, oppure all'elaborazione di informazioni aggiuntive rispetto a quelle normalmente fornite dai dispositivi diagnostici tradizionali.

Un caso d'uso è quello di RiAtlas Healthcare[85], dispositivo medico basato su IA che consente di seguire la deospedalizzazione del paziente. È un software standalone accessibile da web che consente di raccogliere i dati del paziente tramite un'interfaccia grafica e di seguire l'andamento del paziente nella fase successiva alla dimissione grazie alla visualizzazione di dati forniti in tempo reale dal paziente tramite smartphone e wearable compatibili. Questa piattaforma consente inoltre di ricevere suggerimenti per la proposta dei codici ICF (Classificazione internazionale del funzionamento, della disabilità e della salute, codici sviluppati per consentire un facile monitoraggio delle condizioni di salute del paziente) più adeguati. I codici ICF sono scarsamente utilizzati a causa della loro numerosità e della conseguente difficoltà a individuare velocemente quelli corretti da assegnare al paziente. Il software riesce invece a fornire suggerimenti adeguati, facilitando la compilazione dei dati della cartella clinica e il monitoraggio delle condizioni di salute del paziente.

Il secondo caso, ipotetico, è costituito da un dispositivo medico per l'ottimizzazione di un piano radioterapico. Esso prende in input le immagini realizzate dalla TC, le immagini di una telecamera esterna e il piano terapeutico del medico e permette il controllo dei dispositivi di radioterapia. L'IA consente di automatizzare la fase di allineamento della macchina e può, grazie anche alla velocità di calcolo e di esecuzione, aggiustare in tempo reale eventuali variazioni del posizionamento del paziente, includendo la correzione di fenomeni fisiologici come gli spostamenti dati dalla respirazione. In fase di analisi, l'IA consente di identificare facilmente strutture anatomiche e punti di repere in modo da registrare facilmente le immagini diagnostiche con le immagini di pianificazione.

Infine ci sono sperimentazioni (incrementate nel periodo COVID) per affiancare sistemi di riconoscimento automatico delle immagini ai sistemi tradizionali: per esempio un algoritmo indica la probabilità che una radiografia polmonare abbia una diagnosi COVID e fornisce al radiologo una priorità per refertare. In questo modo i pazienti infettivi restano meno tempo in sala d'attesa e diminuisce il rischio che il virus si diffonda. Altri algoritmi individuano particolari anomali e li evidenziano al personale sanitario.

[85] https://www.riatlas.it/index.html

22.2 Sistemi informatici ospedalieri e regionali

Nel caso di sistemi, quali quelli ospedalieri (per esempio, ottimizzazione logistica dei reparti, gestione automatica del magazzino) e regionali (per esempio, sistemi epidemiologici, comprensivi di previsione della spesa), fascicolo sanitario elettronico del cittadino (che consolida non solo il risultato delle attività ospedaliere e ambulatoriali ma anche di piani diagnostico-terapeutico-assistenziali, protesica, vaccinazioni, malattie infettive), sistemi prescrittivi (la cosiddetta prescrizione elettronica) e sistemi di prenotazione (CUP regionali), non si ha evidenza di applicazioni significative dell'IA se non a titolo sperimentale.

Un software dotato di IA potrebbe suggerire l'organizzazione dei turni, delle sale operatorie e dei magazzini ottimizzando le risorse. Il software riceverebbe in input i dati storici dell'azienda ospedaliera, creando un database rappresentativo del flusso e delle attività tipiche. Dopo un adeguato tempo di raccolta dati, il software potrebbe fornire previsioni dei valori futuri dei parametri, individuando i momenti e i reparti che richiedono maggiori risorse umane e materiali, e organizzando materiali e turni, consentendo idealmente di fornire un servizio di cura migliore e più rapido.

https://www.youtube.com/watch?v=ZPXCF5e1_HI

"How It Works: IBM Watson Health"

Spiegazione di come funziona IBM Watson in healthcare.

(durata: 4 minuti e 3 secondi)

https://www.youtube.com/watch?v=NFUZBT0HKRg

"Health ready for AI?"

Interventi sugli sviluppi della telemedicina.

(durata: 23 minuti e 43 secondi)

23 Settore militare

È diffusa la convinzione che i campi di battaglia del futuro, quando l'IA si sarà compiutamente sviluppata, non vedranno più combattenti affrontarsi in campo aperto, ma sistemi intelligenti operare in remoto su una molteplicità di livelli e di domini di conflitto.

Nei paragrafi successivi sono approfonditi i seguenti utilizzi dell'IA in ambito militare:

- raccolta ed elaborazione dei dati di intelligence;
- controllo dei veicoli autonomi (capitolo 24.1).

Si sta anche prevedendo di usare l'IA nei teatri di guerra convenzionali in modo che aiuti nell'identificazione dei bersagli e nel puntamento delle armi, lasciando agli umani la decisione se fare fuoco o meno. Altri impieghi dell'IA in battaglia sono ipotizzati negli scontri aerei (vedere paragrafo 24.1.2) e navali[86].

Ulteriori notizie sulle evoluzioni dell'IA in ambito militare:

- un algoritmo di IA ha sconfitto un pilota di caccia F-16 statunitense in una simulazione di un combattimento aereo[87];
- un LMS (learning management system) ottimizzato attraverso l'IA permette di creare un processo di addestramento dei piloti che si adatta e migliora continuamente[88];
- è in corso un progetto di sviluppo di un prototipo di un centro di wargaming in cui, al fine di fornire maggiori metriche e una migliore formazione[89], saranno integrate tecnologie avanzate tra cui: apprendimento automatico, teoria dei giochi, modellazione e simulazione multidominio e analisi predittiva dei dati;
- è in corso l'implementazione di soluzioni di gestione dei rischi della supply chain alimentate dall'intelligenza artificiale e screening di fornitori per conto del Dipartimento della Difesa (DoD)[90];
- per i nuovi satelliti militari, si prevede di utilizzare la tecnologia IA per consentire non solo la produzione e gli aggiornamenti software nello spazio, ma anche la rilevazione di anomalie hardware e la gestione delle loro riparazioni[91];
- sono in corso studi per l'utilizzo dell'IA per addestrare e migliorare l'accuratezza degli algoritmi di visione per la navigazione dei velivoli dell'Air Force USA[92].

[86] https://warontherocks.com/2018/06/the-dawn-of-artificial-intelligence-in-naval-warfare/.
[87] https://militaryembedded.com/ai/machine-learning/darpa-to-hold-aerial-combat-simulation-between-f-16-pilots-and-ai.
[88] https://militaryembedded.com/avionics/safety-certification/ai-powered-pilot-training-initiative-supported-by-cae-usa.
[89] https://militaryembedded.com/ai/big-data/wargaming-center-prototype-in-development-for-usmc.
[90] https://militaryembedded.com/ai/deep-learning/ai-technology-to-be-leveraged-for-dod-supply-chain-risk-management.
[91] https://militaryembedded.com/comms/satellites/satellites-to-be-developed-with-ai-to-enable-fast-software-upgrades.
[92] https://militaryembedded.com/ai/big-data/ai-algorithms-to-improve-air-force-navigation-capabilities.

Gli ampi investimenti in ambito militare in corso presso alcuni Paesi hanno anche fatto nascere discussioni sull'applicabilità dell'intelligenza artificiale all'interno di un'eventuale risposta nucleare[93].

23.1 Raccolta ed elaborazione dei dati di intelligence

Relativamente all'intelligence, la sfida è quella di acquisire i dati in ambienti contesi e trasmetterli ed elaborarli in modo efficiente. Questo compito è reso più difficile dal volume ed eterogeneità dei dati stessi e dal fatto che il conflitto si verifica spesso in aree a cui sono stati negati i dati, rendendo quasi impossibile per questi sistemi accedere alle informazioni che i militari vogliono o non sanno ancora di aver bisogno. I sistemi di IA in questo ambito hanno l'intento di fornire ai combattenti informazioni mission-critical e tempestive al limite tattico.

L'intelligence militare fa uso di fonti testuali, che si tratti di messaggi di smartphone, e-mail, post sui social media, documenti e simili. L'analisi di questi dati per parole chiave, comportamento di gruppo e intelligenza utilizzabile è un'attività che spesso va oltre un analista umano e pertanto necessita il supporto di sistemi di IA..

23.2 Veicoli autonomi militari

I veicoli autonomi rappresentano la risposta tecnologica più avanzata per ridurre le perdite umane e combattere in modo vincente il nemico, ottimizzando i tempi e metodi per pervenire all'obiettivo del combattimento. Per maggiori dettagli, vedere il capitolo 28.

[93] https://www.notizie.ai/possiamo-lasciare-allintelligenza-artificiale-il-controllo-della-risposta-nucleare/.

24 Strade intelligenti

L'ottimizzazione del traffico, il cui impatto era stimato, nel 2018, dal World Economic Forum intorno ai 87 B$ nei soli USA[94], rappresenta per l'IA un campo di utilizzo privilegiato grazie alla possibilità, indotta da sensori e telecamere, di raccogliere dati e immagini per costruire modelli di previsione da applicare alla gestione delle infrastrutture di trasporto.

Tra i principali casi d'uso di tecniche di *image & object detection* troviamo[95] l'identificazione dei flussi di traffico agli incroci e il tracciamento di uno specifico veicolo al fine di ricostruire i suoi movimenti e avere un modello attendibile di simulazione dei flussi dei veicoli, facendo anche una classificazione per tipologia (auto, autobus, autocarri, motocicli) [96].

La capacità di rilevare i volumi e i modelli di traffico permette di avvisare i conducenti delle condizioni stradali, rilevare quando i veicoli lasciano la strada o sono coinvolti in incidenti, avvisare i servizi di emergenza e le autorità appropriate, fornire agli automobilisti indicazioni di deviazione dei flussi di traffico nel caso di incidenti e suggerire loro traiettorie alternative e interventi sulle velocità per evitare l'eccesso di traffico. Questo contribuisce a rendere le condizioni stradali più sicure, come si prefigge il piano Smart Road di Anas[97], che prevede 3.000 km di strade intelligenti entro il 2030, integrando anche sistemi di rilevazione del meteo.

Applicazioni dell'IA deputate al monitoraggio del traffico, estremamente promettenti, sono quelle legate al controllo del traffico stesso attraverso l'ottimizzazione dei tempi di cambio di stato dei semafori[98] che, in funzione del traffico misurato, permettono di minimizzarne la congestione complessiva.

L'addestramento è di tipo reinforcement learning: non è necessario un dataset iniziale, ma un simulatore dell'ambiente in cui l'applicazione opererà. La disponibilità di progetti open source come il Simulation of urban mobility (SUMO)[99] offre l'ambiente dove modellare la propria rete autostradale e dove gli agenti apprendono attraverso azioni e feedback che ne definiscono il comportamento appropriato al fine di ottenere un premio: la più bassa congestione del traffico.

È anche possibile addestrare i sistemi con dataset derivati da quelli creati per l'addestramento delle auto a guida autonoma, come il KITTI Vision Benchmark Suite[100] o lo Stanford Cars Dataset[101], solo per citarne alcuni, e dalla possibilità di generare immagini sintetiche attraverso tecniche di raytracing su modelli di auto e che permettono di superare le difficoltà

[94] https://www.weforum.org/agenda/2019/03/traffic-congestion-cost-the-us-economy-nearly-87-billion-in-2018/.
[95] https://www.aicitychallenge.org.
[96] https://mobility.here.com/learn/smart-transportation/traffic-ai-real-life-use-case#:~:text=The%20term%20traffic%20AI%20refers,them%20to%20the%20traffic%20infrastructure.
[97] ANAS Smart Road - https://www.stradeanas.it/it/smart-road-anas.
[98] https://arxiv.org/pdf/1803.11115.pdf.
[99] https://sumo.dlr.de/docs/index.html.
[100] http://www.cvlibs.net/datasets/kitti/.
[101] http://ai.stanford.edu/~jkrause/cars/car_dataset.html.

degli algoritmi ad apprendimento supervisionato (paragrafo 3.2) basati sulla disponibilità di un numero adeguato di immagini già classificate[102].

[102] https://ieeexplore.ieee.org/document/8954694.

25 Smart city

L'IA può essere applicata al contesto evolutivo urbanistico identificato dalle cosiddette "smart city". Per raggiungere questo obiettivo è necessario analizzare dati raccolti dalle piattaforme IoT, realizzate in alcuni casi dagli stessi attori che hanno installato altre infrastrutture nelle città. L'IA, e in particolare il *machine learning*, permettono di utilizzare in tempo reale i dati raccolti, abilitando azioni che aiutino i cittadini ad avere una migliore esperienza all'interno dei centri urbani, generando vantaggi sia ai singoli sia alla comunità.

A Barcellona, contatori d'acqua intelligenti hanno aiutato la città a risparmiare più di 50 milioni di euro all'anno. In Corea del Sud, una città ha ridotto i costi operativi degli edifici del 30% dopo aver attivato sensori intelligenti per regolare l'utilizzo di acqua ed elettricità[103].

Di seguito vengono riportati altri ambiti, attuali e futuribili, in cui l'intelligenza artificiale può essere utilizzata per migliorare i tempi di risposta e rendere le nostre città ancora più smart[104].

- *Controllo del traffico,* come già illustrato nel capitolo precedente.
- *Controllo della temperatura negli stabili*, connesso con il controllo dell'energia, oggetto del prossimo capitolo.
- *Gestione dei parcheggi*: l'utilizzo di sensori, insieme alla geolocalizzazione dei cittadini, potrebbe permettere di identificare molto più facilmente i parcheggi liberi nella zona di interesse.
- *Dimensionamento dei mezzi pubblici disponibili*: l'utilizzo di applicazioni di IA che, tramite l'acquisizione di immagini, video e dati di geolocalizzazione degli utenti, permetterebbero di gestire i mezzi disponibili non in base ad una pianificazione standard ma in base ai reali flussi dei passeggeri nell'arco della giornata.
- *Monitoraggio edifici ed infrastrutture*, come sarà illustrato ai capitoli 26 e 27.
- *Sicurezza urbana,* per cui si rimanda al paragrafo 45.2.

Da questo elenco si evincono chiaramente alcuni problemi legati alla sicurezza ed alla privacy delle persone.

https://www.youtube.com/watch?v=Xir-SeUK6zs

"Smart City Powered by Artificial Intelligence "

Trasformazione di qualsiasi città in un ambiente intelligente e sicuro con le 14 funzioni di sicurezza e analisi di IronYun su un unico dispositivo scalabile e potente: AI NVR.

(durata: 6 minuti circa)

https://www.youtube.com/watch?v=NjGDIQ6GV2c

"Dubai nel 2030"

Meet Frey, the Smart City AI of Dubai 2030.

(durata: 1 minuto e 07 secondi)

[103] https://newsroom.cisco.com/feature-content?type=webcontent&articleId=677558.
[104] https://www.ai4business.it/news/smart-city-intelligenza-artificiale/.

26 Smart building e risparmio energetico

Il Governo italiano scrive: "L'Italia intende perseguire un obiettivo indicativo di riduzione dei consumi al 2030 pari al 43% dell'energia primaria e al 39,7% dell'energia finale rispetto allo scenario di riferimento PRIMES 2007"[105].

Il patrimonio immobiliare italiano è vecchio di 40 anni e questo non depone a favore del paradigma *smart building (*o *building 4.0)*, secondo il quale tutto l'edificio viene progettato sin dal principio per essere intelligente e connesso con il mondo esterno. La sfida è quindi rendere gli edifici, ormai obsoleti, intelligenti.

Per usare razionalmente l'energia, l'Europa ha definito un indicatore chiamato SRI cioè *smart readiness indicator*[106]. Questo indicatore deve basarsi fra le altre cose su: "la capacità di adattare la propria modalità di funzionamento in risposta alle esigenze dell'occupante, prestando la dovuta attenzione alla facilità d'uso, al mantenimento di condizioni di benessere igrotermico degli ambienti interni e alla capacità di comunicare dati sull'uso dell'energia".

Per avere un edificio intelligente serve un sistema di automazione interconnesso, i cosiddetti BACS (*building automation and control system*). Il funzionamento dei BACS, base anche dell'SRI, si fonda sul concetto per cui l'energia di un edificio deve essere prodotta e distribuita a seconda della richiesta e non, come avviene oggi, su congetture e supposizioni oramai obsolete come gli orari di funzionamento delle centrali in base alle fasce climatiche (anch'esse ormai obsolete a causa del riscaldamento globale).

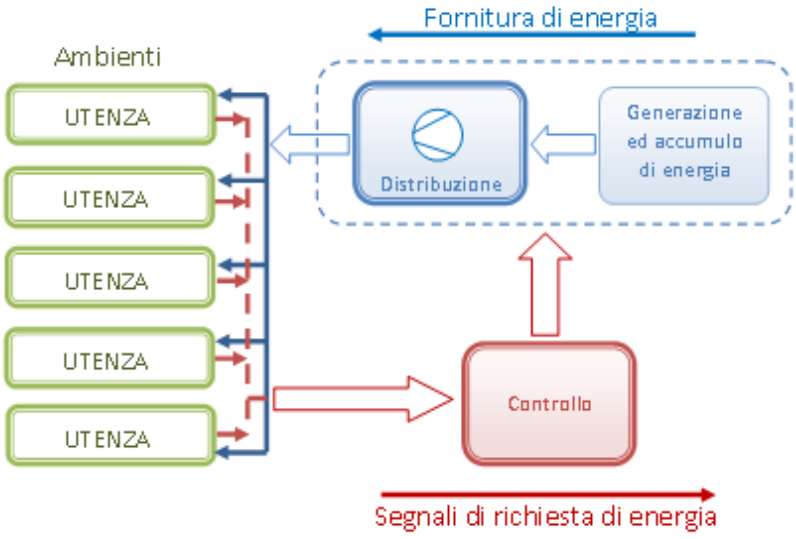


Figura 15 - Schema dell'energia richiesta e fornita nel caso di un impianto tecnico[107]

[105] PNIEC, Piano Nazionale Integrato per l'Energia e il Clima 2030, MiSE 21.01.2020: https://www.mise.gov.it/index.php/it/energia/energia-e-clima-2030
[106] https://smartreadinessindicator.eu/.
[107] Fonte: AiCARR - ANIE CSI, "Vademecum BACS: Guida all'impiego dei sistemi di automazione, controllo e gestione tecnica degli edifici alla luce della norma UNI EN 15232-1:2017".

Una volta abilitato l'edificio 4.0 (con sensori e tecnologie di tipo edge computing) e interconnesso, entrano in gioco i cosiddetti big data perché bisogna elaborare grandezze estremamente difficili da prevedere come, per esempio:

1. il clima previsto all'esterno dell'edificio in termini di temperatura ed irraggiamento, tenendo conto degli ombreggiamenti;
2. l'uso previsto degli impianti di climatizzazione in ogni singolo locale; esso dipende dall'occupazione e da altri fattori che influenzano il clima interno come illuminazione, irraggiamento, tipologia di occupante e di attività del singolo locale;
3. il comportamento atteso degli utenti.

Per alcuni gestori, i dati riguardano migliaia di edifici di tipi diversi: villa unifamiliare, condominio, edifici terziari, uffici, hotel, centri commerciali, edifici industriali, ecc. Per rendere questi dati uniformi, e quindi elaborabili da algoritmi di IA per gestire in modo ottimale e sistematico problemi con numerose variabili, è stato avviato il progetto Haystack.[108]

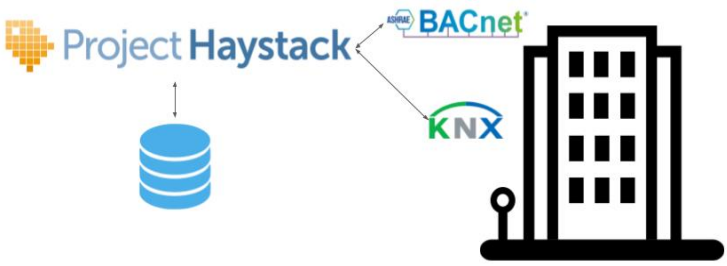


Figura 16 – Progetto Haystack

https://youtu.be/eVJi9pzF8yI

"Today's Intelligent buildings with Project Haystack"

In questo video si illustra il progetto Haystack.

(durata: 15 minuti e 45 secondi)

[108] https://project-haystack.org/

27 Infrastrutture critiche

L'obsolescenza delle infrastrutture critiche è un tema di enorme importanza. Per esempio, la maggior parte dei ponti è stata costruita decenni fa e in molti casi hanno superato la durata per la quale era stata progettata; nei soli Stati Uniti, secondo un report del 2017, ci sono ben 56.000 ponti con problemi che richiederebbero 123 miliardi di dollari per la loro ristrutturazione[109]. Più in generale, McKinsey stima l'impatto dell'intelligenza artificiale applicato al problema dell'*anomaly detection* in tutte le industrie tra 1 e 1,4 T dollari[110].

L'IA può essere utilizzata per analizzare i dati provenienti dai sensori utilizzati per monitorare lo stato di edifici e infrastrutture e identificare andamenti potenzialmente pericolosi così da innescare degli allarmi tempestivi in caso di eventuali problemi. Il monitoraggio attraverso sensori permette di raccogliere dati con elevata frequenza e di valutare correttamente la velocità di degrado dell'infrastruttura[111].

L'utilizzo di ulteriori apparati intelligenti (p.e. valvole) da attivare in caso di necessità potrebbe permettere di ridurre significativamente gli eventuali danni (p.e. perdite di acqua).

Il rischio dei tradizionali metodi di anomaly detection attraverso sensori è dato dal rumore di fondo. Questo disturbo si presenta nelle rilevazioni dei sensori e tende a nascondere gli *early warnings* che caratterizzano un guasto anche giorni prima della sua manifestazione palese. Il rumore di fondo può anche causare picchi non dovuti ad anomalie, ma che segnalano falsi positivi e che inducono i tecnici a innalzare le soglie di allarme, nascondendo gli early warning significativi. Sofisticati algoritmi di machine learning applicati all'anomaly detection sono invece in grado di apprendere come deve essere un segnale normale, incluso il rumore di fondo, e distinguere misure che escono dal pattern associato a quel sensore.

Un esempio di questi algoritmi è il *multivariate state estimation technique* (MSET)[112], utilizzato per anni nelle centrali nucleari, infrastrutture dove "f*ailure is not an option*". L'algoritmo si compone di due elementi: uno che valuta lo stato del componente monitorato e un altro che stima il guasto potenziale. L'apprendimento avviene a partire da misure prese durante il normale funzionamento dell'apparato ed è, quindi, meno oneroso di un algoritmo ad apprendimento supervisionato.

Il rilevamento delle anomalie con sensori, tradizionalmente, va affiancato dall'ispezione visiva condotta da un esperto addestrato e che comporta un dispendio di tempo notevole, è laboriosa, costosa e spesso pericolosa. Oggi è notevolmente semplificata dall'uso di telecamere remote, anche in congiunzione con veicoli aerei autonomi (unmanned aerial vehicles, UAV), e dagli algoritmi di computer vision basati sul deep learning.

---

[109] "ASCE's 2017 infrastructure report card: bridges". Reston: American Society of Civil Engineers; 2017. https://www.infrastructurereportcard.org/cat-item/bridges/.
[110] https://www.mckinsey.com/featured-insights/artificial-intelligence/visualizing-the-uses-and-potential-impact-of-ai-and-other-analytics.
[111] https://www.sciencedirect.com/science/article/pii/S2095809918308130#b0005
[112] https://www.ne.anl.gov/codes/mset/

La computer vision permette di automatizzare il processo di *damage detection*, come le crepe nelle strutture in acciaio dei ponti[113], le crepe nell'asfalto[114] o nel cemento[115], con un'alta accuratezza. In entrambi i casi si usano modelli di classificazione delle immagini basati sulle reti neurali convoluzionali (*convolutional neural network*, CNN).

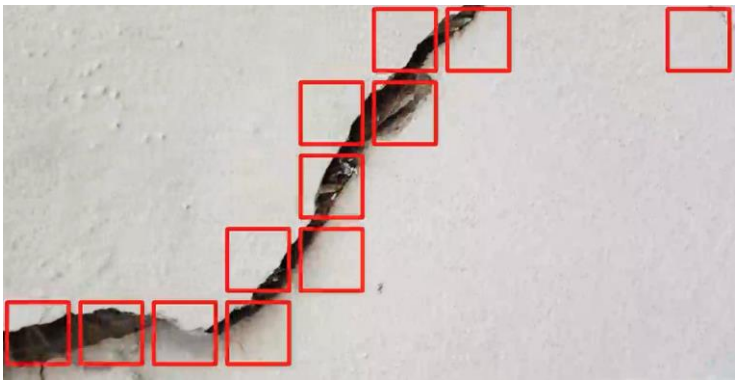


Figura 17 - Analisi a griglia per rilevamento crepe con reti neurali CNN[116]

Ottimi risultati sono stati ottenuti con tecniche di *object detection*, basati sulle R-CNN (region based convolutional neural networks), dove si cerca di individuare all'interno dell'immagine i danni inserendoli in apposite categorie. Applicandole al riconoscimento di crepe nel cemento, corrosione dei bulloni e delaminazione dell'acciaio[117], si è raggiunta una precisione media dell'87,8%.

Un'altra interessante applicazione è l'identificazione e mappatura automatica delle buche stradali. A tal fine, l'università tedesca di Ilmenau ha messo a disposizione un dataset di addestramento denominato GAP[118], utilizzato da oltre 125 team di ricerca nel mondo, per la costruzione di modelli di classificazione della qualità del manto stradale seguendo le direttive della *German Road and Transportation Research Association* (FGSV) secondo un approccio per la collezione di dati denominato RMA, road monitoring and assessment.

[113] https://journals.sagepub.com/doi/abs/10.1177/1475921718764873.
[114] https://ieeexplore.ieee.org/document/7533052.
[115] https://onlinelibrary.wiley.com/doi/abs/10.1111/mice.12263.
[116] Immagine degli autori.
[117] https://onlinelibrary.wiley.com/doi/abs/10.1111/mice.12334.
[118] https://www.tu-ilmenau.de/en/neurob/data-sets-code/gaps/.

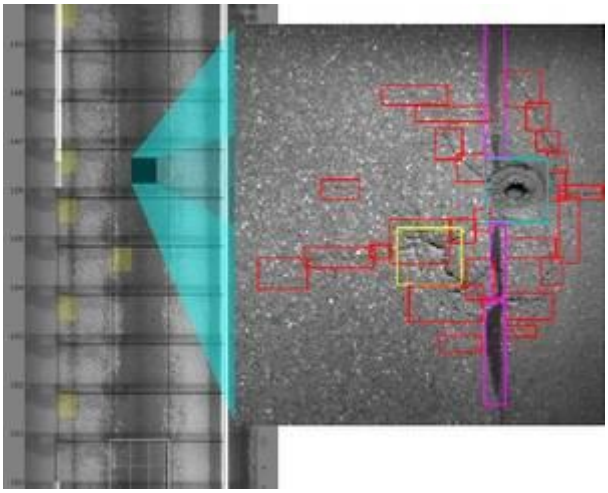


Figura 18 - Mappatura del manto stradale[119]

Un altro esempio di applicazione di tecniche di computer vision e IA riguarda la gestione del rischio connesso ai disastri naturali. Si tratta del Open Cities AI Challenge[120], un concorso lanciato da un dipartimento della Banca Mondiale, il Global Facility for Disaster Reduction and Recovery (GFDRR), per selezionare tecniche di identificazione automatica delle infrastrutture basate su rilevamenti aerei fatti da droni. Queste sono necessarie alle aree urbane a forte crescita di popolazione ed esposte a rischio di esondazione, erosione, terremoti e tempeste costiere che richiedono una costante mappatura del territorio. Il GFDRR ha messo a disposizione un dataset di immagini annotate a mano di aree urbane africane per addestrare gli algoritmi partecipanti al concorso.

Figura 19 – Esempio di mappatura Dar es Salaam (Tanzania)[121]

119 https://www.tu-ilmenau.de/en/neurob/data-sets-code/gaps/.

120 https://towardsdatascience.com/the-open-cities-ai-challenge-3d0b35a721cc; https://www.drivendata.org/competitions/60/building-segmentation-disaster-resilience/page/151/.

121 https://towardsdatascience.com/the-open-cities-ai-challenge-3d0b35a721cc.

Nel 2018 era stato lanciato un simile concorso anche nell'ambito dell'AICrowd[122], organizzazione crowdsourcing in ambito IA. È stato promosso da un'organizzazione no-profit operante in 60 nazioni, la Humanity & Inclusion, che ha un dipartimento che si occupa di emergenze umanitarie e interviene in caso di disastri civili e naturali. In questo caso, l'utilizzo del deep learning applicato alla computer vision ha l'obiettivo di ricostruire, a seguito di un disastro, le mappe territoriali. In caso di disastro, infatti, è fondamentale avere informazioni su strade non percorribili, entità dei danni alle abitazioni civili o movimenti di profughi, e le immagini possono essere reperite rapidamente da varie fonti, come nano-satelliti, droni e satelliti convenzionali a elevate altitudini. Algoritmi di segmentazione automatica delle immagini possono identificare rapidamente edifici intatti, strade e infrastrutture critiche per interventi efficaci e tempestivi.

[122] https://www.aicrowd.com/challenges/mapping-challenge.

28 Veicoli autonomi

I veicoli autonomi sono presentati in due famiglie: militari e civili.

Per valutare in maniera appropriata il grado di autonomia decisionale di un veicolo è necessario disporre di un criterio di classificazione il più possibile rigoroso: attualmente la rappresentazione migliore del livello di autonomia (LOA, *levels of autonomy*) è fornito dalla Sheridan autonomy scale, una scala da 1 a 10 che mette in relazione l'uso del computer di bordo con l'interazione dell'essere umano:

1. il computer non offre alcuna assistenza; l'essere umano fa tutto;
2. il computer offre un set completo di alternative d'azione;
3. il computer restringe la selezione verso il basso per alcune scelte;
4. il computer suggerisce una singola azione;
5. il computer esegue un'azione, se l'essere umano l'approva preventivamente;
6. il computer consente un tempo limitato all'essere umano di porre il veto prima dell'esecuzione automatica di un'azione;
7. il computer esegue automaticamente; quindi ne informa l'essere umano;
8. il computer informa l'uomo dopo l'esecuzione automatica, solo su richiesta dell'essere umano;
9. il computer informa l'essere umano dopo l'esecuzione automatica, solo se decide di farlo;
10. il computer decide tutto e agisce in modo autonomo, ignorando l'essere umano.

I veicoli autonomi sono in uso ormai da vari anni in varie aree del mondo. In particolare quelli militari sono stati sviluppati e usati soprattutto in aree "calde" (come Israele), ma anche in USA, Cina ed anche in Italia. I veicoli autonomi sono attualmente utilizzati anche per scopi civili, come la sorveglianza fisica di infrastrutture critiche (p.e. aeroporti, centrali elettriche e impianti petroliferi).

28.1 Veicoli autonomi militari

28.1.1 Unmanned ground vehicle (UGV)

L'UGV, autonomo, piccolo, leggero, di basso profilo, silenzioso, agile su tutti i terreni e facile da controllare, segue il soldato ed è utilizzato per il pattugliamento e la ricognizione, come piattaforma per il trasporto di armi pesanti o di soldati feriti e come postazione di osservazione.

Con l'incremento dei dati raccolti, degli scenari, delle condizioni e variabili di stato del sistema, gli UGV possono operare in forma autonoma in ambienti operativi assai diversi ed ottimizzare le azioni di supporto logistico, di attacco e di difesa.

Figura 20 - UGV per uso militare logistico

Figura 21 - UGV per uso militare operativo offensivo-difensivo

In Figura 21[123] è rappresentato un UGV dotato di telecamere ottiche mobili con autofocus, telecamere infrarossi termiche, sensori di varie tipologie a seconda della destinazione di uso del mezzo. Gli UGV utilizzati nel controllo perimetrale di centrali di produzione energia atomiche effettuano il rilevamento delle radiazioni.

Figura 22 – UGV di sicurezza

In Figura 23 è rappresentato un UGV dotato di un'unità CAFS (*compressed air foam system*) che fornisce capacità antincendio.

[123] Fonte per le figure di questo paragrafo: amstaf-ugv.com.


Figura 23 - UGV antincendio e particolare del cannone di erogazione con telecamera per puntamento

In Figura 24 è rappresentato un UGV destinato alle forze dell'ordine per il controllo di tumulti e rivolte. Tale sistema è stato sponsorizzato dalla NATO ed inserito nel programma per le armi "non letali" dal Dipartimento della difesa degli Stati Uniti e dal Dipartimento della difesa nazionale del Canada. Nello specifico è dotato di sistema per il lancio di lacrimogeni che può indirizzare i lanci usando un sofisticato sistema di visione in grado di processare le immagini acquisti da telecamere e segnali catturati da sensori.

Figura 24 - UGV per repressione rivolte

In Figura 25 è rappresentato un tipico esempio di UGV di tipo militare dotato di strumentazione offensiva con mitragliatore da campo, puntatore controllato attraverso segnali provenienti dall'acquisizione e elaborazione di immagini da telecamere per visione diurna, notturna e termica.

Figura 25 - UGV con mitragliatore da campo

28.1.2 Unmanned Aerial Vehicles (UAV)

La storia dei veicoli aerei militari senza pilota può essere fatta risalire già alla metà del 1800 con l'attacco degli austriaci a Venezia mediante palloni aerostatici caricati di esplosivo. Qualche esperimento significativo è stato effettuato durante la Seconda guerra mondiale. Negli anni '60 del secolo scorso gli studi sui veicoli senza pilota si sono sviluppati massimamente a livello globale.

Nell'ultimo ventennio gli UAV militari strategici, inclusi quelli a controllo remoto (ROA – remotely operated aircraft e RPA – remotely piloted aircraft), hanno avuto un enorme sviluppo dal punto di vista tecnologico e operativo: ormai non esistono scenari di guerra classica, di polizia internazionale o altri ambiti "non convenzionali" che non utilizzino diffusamente questi veicoli.

Attualmente le missioni possibili degli UAV sono molteplici: si può andare dalla missione investigativa di ricognizione alla missione armata di attacco su obiettivi precisi.

Ovviamente, insieme a queste tecnologie IA si stanno discutendo criteri etici, di certificazione e di convalida per questi sistemi autonomi.

Figura 26 - MQ-9A Reaper (Reaper B) dell'Aeronautica militare italiana[124]

28.2 Veicoli autonomi civili

I veicoli civili unmanned sono distinti in più famiglie, oggetto dei paragrafi che seguono.

28.2.1 Settore industriale

Il gruppo automotive BMW usa robot intelligenti in logistica e applicazioni IA per ottimizzare il flusso dei materiali. In un primo progetto pilota, BMW ha equipaggiato robot logistici e robot di trasporto intelligenti (Figura 27) di speciali moduli IA per il loro coordinamento e per fornire loro la capacità di riconoscere persone e oggetti e di identificare percorsi alternativi.

Figura 27 - Robot in fabbrica BMW

[124] General Atomics https://www.ga-asi.com/remotely-piloted-aircraft/mq-9a.

28.2.2 Settore agrario

Nella storia agraria, fin dagli antichi romani, l'occhio umano è stato quasi l'unico "sensore" che ha guidato la viticoltura fornendo una valutazione soggettiva su resa, crescita vegetativa e stato della pianta. Le tecniche moderne richiedono un monitoraggio oggettivo e continuo sulla misura di parametri chiave per il processo decisionale.

È stato sviluppato il progetto di un UGV denominato VineRobot**.** È dotato di intelligenza artificiale con tecnologie di rilevamento non invasive quali la fluorescenza per misure su clorofilla, la visione artificiale red green blue (RGB) e termografia per monitorare le piante e grappoli mentre si muove autonomamente tra i filari. Tra i parametri misurati vi sono: la resa e composizione dell'uva, la crescita vegetativa, lo stato dell'acqua.

Figura 28 - VineRobot in fase operativa

28.2.3 Veicoli aerei

Gli UAV sono in parte usati per la consegna di pacchi (paragrafi 18.2 e 18.3). Fino a ora, i carichi trasportabili sono limitati, ma gli UAV potrebbero fornire un importante soluzione ai centri urbani, eliminando il traffico dalle strade e portandolo in aria.

Figura 29 - UAV in fase di sperimentazione presso DHL[125]

i

28.2.4 Veicoli di superficie

Figura 30 - USV (Unmanned surface vehicles)

In Figura 30 sono rappresentate due configurazioni di USV (*unmanned surface vehicles*).

Possono essere usati per batimetria (per esempio, verifica e controllo delle acque da agenti inquinanti, prelievi in laghi, canali o comunque aree acquatiche non praticabili con normali imbarcazioni), come basi di decollo e atterraggio per droni o, se dotati di telecamere, per la sicurezza e videosorveglianza di bacini idrici.

125 https://www.dhl.com/content/dam/downloads/g0/about_us/logistics_insights/DHL_TrendReport_UAV.pdf.

- QUARTA PARTE: NORMATIVA E ETICA

29 Normativa italiana ed europea

L'intelligenza artificiale, come tutte le nuove tecnologie, rappresenta una grande opportunità per l'umanità. Essa, per esplicare al meglio i suoi effetti e le sue potenzialità, necessita però di essere adeguatamente regolamentata.

Ad oggi non sono ancora state approvate, in via definitiva, normative che disciplinano l'utilizzo dell'IA: più precisamente il dibattito giuridico attiene alla necessità o meno di emanare specifiche discipline ad hoc per l'IA (ed eventualmente in quali ambiti) oppure se invece è sufficiente applicare a tali tecnologie le prescrizioni del quadro normativo già oggi vigente.

Uno dei temi cardine è quello relativo alla definizione giuridica di IA. Come già evidenziato al capitolo 5, non esiste un'unica definizione di intelligenza artificiale a livello scientifico e questo impatta fortemente in ambito giuridico, perché non consente di delineare l'ambito di applicazione delle norme.

A livello europeo, in particolare, sta crescendo la consapevolezza della necessità di una regolamentazione specifica sull'intelligenza artificiale: prova ne sono i vari pareri degli esperti e soprattutto le linee guida per una "intelligenza artificiale affidabile" emanate dalla Commissione europea[126].

Un ulteriore documento, sempre della Commissione europea del 2019[127], sottolineava il vuoto normativo e i profili sui quali l'IA avrà ripercussioni, evidenziando, in particolare, la necessità di disciplinare i seguenti aspetti:

- intervento dell'elemento umano e supervisione delle applicazioni;
- robustezza tecnica e sicurezza negli ambienti di lavoro;
- protezione e governo dei dati;
- trasparenza dei sistemi IA;
- diversità, non discriminazione ed equità;
- benessere della società e dell'ambiente;
- responsabilità.

[126] Independent high-level expert group on artificial intelligence. *Policy and investment recommendations for trustworthy AI*. Bruxelles: European Commission, 2019. https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence.

[127] *White Paper on Artificial Intelligence: a European approach to excellence and trust.* Bruxelles: European Commission, 2019. https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

Il Parlamento europeo ha raccolto il testimone con un'articolata relazione[128], recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale, che include una proposta di Regolamento "sui principi etici per lo sviluppo, la diffusione e l'utilizzo dell'intelligenza artificiale, della robotica e delle tecnologie correlate". La Risoluzione contiene in allegato una proposta di Regolamento comunitario nel quale, oltre a presentare una prima definizione giuridica di IA[129], il Parlamento UE indica i principi cardine che devono essere rispettati dalla IA (eticità, antropocentrismo, sicurezza, trasparenza, assenza di discriminazioni, responsabilità sociale, sostenibilità ambientale, rispetto del trattamento dei dati), prevedendo altresì un certificato europeo di conformità etica. La proposta di regolamento dovrà attraversare tutto l'iter legislativo per essere approvata.

Relativamente poi agli standard di produzione, si è registrato negli ultimi anni un grande aumento dell'interesse e delle attività intorno all'IA da cui emerge la necessità di sviluppare un quadro coerente a livello internazionale. In risposta a questa esigenza, ISO e IEC hanno creato un comitato di standardizzazione sull'IA, l'ISO/IEC JTC 1 SC 42, mentre il gruppo dedicato alla privacy (ISO/IEC JTC 1 SC 27 WG 5) ha appena avviato i lavori sulla materia. L'IEEE è anch'essa molto attiva, in particolare nel campo dell'etica dei sistemi autonomi e intelligenti.

La Commissione europea ha anch'essa posto in atto iniziative specifiche sulla standardizzazione in materia di IA, chiarite nel Rolling Plan for ICT Standardisation[130], documento emesso annualmente dalla Commissione europea.

Intervista ad Alessandro Curioni, Presidente Di.Gi. Academy

Classe 1967 è sposato e padre di due figli. Nasce giornalista e fonda la sua prima azienda nel 1989 con il giornalista e scrittore Giorgio Cajati: Fast-Press s.r.l. Dal 1995 inizia a interessarsi alle nuove tecnologie e nel 2003 pubblica per Jackson Libri il volume "Hacker@tack" dedicato alla sicurezza informatica. Da questa esperienza, e dopo sette anni di consulenza nel settore della cyber security, nasce, nel 2008, DI.GI. Academy, azienda specializzata nella formazione e nella consulenza, della quale è ancora oggi Presidente e per la quale svolge attività di consulenza e formazione in primarie aziende nazionali e internazionali. A metà del 2014, nell'ambito della sua passione per i giochi, fonda, con altri tre appassionati, Pendragon Game Studio, dedicata all'ideazione e pubblicazione di giochi da tavolo. Nel 2015 riprende la sua

[128] Proposta di risoluzione recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate (2020/2012(INL), disponibile presso: https://www.europarl.europa.eu/doceo/document/A-9-2020-0186_IT.html#title1.

[129] art. 4 Proposta di regolamento: Ai fini del presente regolamento si intende per: a) "intelligenza artificiale", un sistema basato su software o integrato in dispositivi hardware che mostra un comportamento intelligente, tra l'altro raccogliendo e trattando dati, analizzando e interpretando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere specifici obiettivi

[130] https://joinup.ec.europa.eu/collection/rolling-plan-ict-standardisation/rolling-plan-2020.

attività pubblicistica scrivendo per testate on line e tradizionali e partecipando a trasmissioni radiofoniche e televisive Nel 2016 pubblica il suo nuovo libro "Come pesci nella Rete – Guida per non essere le sardine di Internet" al quale seguono altri quattro libri e un quinto uscito nel 2019 dal titolo "Cyber War". Dal 2018 insegna all'Università Cattolica di Milano "Sicurezza dell'informazione".

Domanda 1. L'IA potrebbe dare un contributo importante a miglioramento della security, ma ci sono soltanto opportunità o anche rischi?

Quasi due anni fa i ricercatori di Open AI, un'organizzazione no profit sostenuta dal fondatore di Tesla Elon Musk, hanno sfornato una nuova intelligenza artificiale specializzata in manipolazione dei testi che ha dimostrato eccezionali abilità tra cui: la generazione massiva di fake news fino alla costruzione di campagne di spamming e phishing molto sofisticate. Su queste premesse si sono rifiutati di rendere pubblico l'intero codice sottostante. Non diversamente suscitò preoccupazione nel 2012 un algoritmo stilometrico per il riconoscimento testuale capace, analizzando 100 mila blog, di identificare in modo automatico l'ottanta per cento degli autori dei post. Se, da un lato, *weak AI* di questo genere potrebbero essere strumenti molto potenti per riconoscere i diversi stili degli autori dei messaggi e quindi proteggere i destinatari dagli attacchi di phishing[131], dall'altro, la tecnologia, come sempre accade nella storia dell'umanità, mostra l'altra faccia della sua medaglia.

Nel specifico potremmo immaginare tre posizioni. Gli apocalittici che evocano lo spettro di un futuro come quello rappresentato nella cinematografia stile Terminator e Matrix. Gli ottimisti a oltranza che, forse sopravvalutando la razionalità umana, pensano che di fronte a macchine tanto potenti l'uomo sarà capace di introdurre dei sistemi di sicurezza, un po' come le leggi della robotica di asimoviana memoria. Infine, ci sono quelli che minimizzano perché in fondo si sta ancora parlando di "*weak AI*", quel particolare tipo di intelligenza artificiale specializzata: simile a un uomo, ma soltanto rispetto a uno specifico compito, in questo caso scrivere. I più pericolosi sono proprio questi ultimi perché si troveranno nel futuro senza accorgersene e questo significa essersi "persi". Non dimentichiamoci che dal primo volo dei fratelli Wright al Concorde sono trascorsi appena 66 anni e quest'anno Internet ne ha compiuti appena cinquantuno.

Aggiungiamo un ulteriore elemento. Per molto tempo le intelligenze artificiali sono state costruite sulla base di schemi di matematica precisione, ovvero chi le sviluppava sapeva esattamente qual era l'obiettivo, quali dati dovevano essere utilizzati per l'apprendimento e soprattutto perché alla fine il sistema avrebbe preso una data decisione.

Le grandi masse di dati disponibili (i big data) hanno reso possibile un addestramento delle intelligenze artificiali di tipo statistico, quindi basato essenzialmente sui dati. Si è compreso che le reti neurali, se hanno abbastanza informazioni, sono in grado di svolgere correttamente il loro compito: sia esso il riconoscimento stilometrico oppure quello di un viso.

[131] https://www.ilsussidiario.net/news/hi-tech/2019/1/4/reati-informatici-intelligenza-artificiale-il-poliziotto-pronto-a-diventare-alleato-dei-ladri/1829795/.

Proprio dal prevalere di un addestramento basato sui big data probabilmente nessun essere umano sarebbe in grado di capire le motivazioni che spingeranno il sistema a fare una scelta piuttosto che un'altra. Possiamo veramente immaginare un domani in cui dovremo confrontarci con "qualcosa" da noi creato che risponde a logiche che ci risultano completamente oscure?

Abbiamo poi un ulteriore rischio. Tecnicamente si chiama "bias", e si presenta quando un sistema mostra dei pregiudizi che lo portano alla discriminazione. Non è difficile immaginare cosa potrebbe accadere se chi sviluppa un'intelligenza artificiale destinata alla prevenzione del crimine fosse razzista. Altrettanto facile è comprendere quali rischi potrebbero derivare se i dati di addestramento fossero semplicemente imperfetti. Per causare un disastro sono sufficienti errori commessi in buona fede che, purtroppo, sono anche i più difficili da prevenire.

Domanda 2. IA e regolamentazione. Se e come sarà possibile governare l'utilizzo delle IA?

Sul tema dell'intelligenza artificiale gli ultimi anni sono stati caratterizzati da due concetti: strategia e linee guida.

Purtroppo si tratta di termini nefasti per due ragioni: da un lato nessuno ha capito bene di cosa si sta parlando (quindi cerchiamo di essere "strategici" e quindi non entriamo nel dettaglio), di conseguenza non si riesce a regolamentare (forniamo delle "linee guida" vincolanti ma non troppo). I risultati finali sono stati molti. L'Unione europea ha pubblicato il suo "Libro Bianco" in cui parla di requisiti obbligatori per applicazioni ad alto rischio, di sicurezza e responsabilità, di governance. Dal Vaticano si è affermato il rischio che l'informazione e la conoscenza possano diventare una esclusiva di grandi gruppi economici. Microsoft e IBM hanno affermato la necessità per gli operatori del settore di confrontarsi sul tema per comprenderne appieno la portata.

Il nostro Governo ha prodotto una strategia per l'intelligenza artificiale che si è posto sulla scia delle prese di posizione in materia dell'OCSE e dell'Unione europea. Il nostro Paese ha sfornato oltre cento pagine e ben 82 raccomandazioni alcune delle quali, dopo la pandemia 2020-2021, mostrano il vero problema della "strategia" per cui vale la pena ricordare Von Clausewitz che a tal proposito notava come "tutto è molto semplice ma non è altrettanto facile".

Prendiamo la scuola. La raccomandazione 11 parla di un corpo docente aggiornato e competente nelle tecnologie digitali e dell'introduzione del coding (basi di programmazione per lo sviluppo del pensiero computazionale, ovvero come "dialogare" con un computer), la 13 di introdurre corsi sull'intelligenza artificiale negli istituti tecnici, la 14 di riprogettare i corsi di laurea nazionali introducendo un numero di crediti formativi adeguati riconducibili ai temi

dell'IA[132]. La tecnologia, anche nelle sue espressione basilari, non sembra ancora essere propriamente nelle corde di una fetta significativa del corpo docente. Aggiungiamo poi che introdurre nuovi insegnamenti nelle scuole superiori non è banale (si dovrebbero anche trovare i docenti), mentre a livello universitario è un'impresa che spesso richiede un iter burocratico di anni.

Tuttavia, detto questo, dietro "linee guida" e "strategie" c'è un elemento di continuità che ricorre almeno a livello europeo e va rilevato. Proprio sul tema degli algoritmi intelligenti, la Commissione europea ha sottolineato l'importanza dell'accountability nel suo documento "Progetto di orientamenti etici per un'IA affidabile". La governance delle IA, come quella di qualsiasi governo voglia definirsi tale, deve partire e non può prescindere dalla responsabilità nel senso più esteso del termine, esattamente il significato che rende perfettamente il termine "accountability".

[132] https://www.ilsussidiario.net/news/intelligenza-artificiale-il-profumo-di-una-rosa-dividera-uomini-e-robot/1938610/.

30 IA e privacy

30.1 GDPR e IA

Come noto, la materia della protezione dei dati personali è oggi disciplinata dal Regolamento UE 2016/679 (GDPR). Il trattamento di dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali è disciplinato invece dalla Direttiva UE 2016/680 e dalle legislazioni nazionali di recepimento.

Il GDPR contiene numerose disposizioni applicabili all'intelligenza artificiale, anche se l'IA non è mai esplicitamente menzionata. È innegabile infatti che il trattamento di dati personali tramite tecnologie di IA può presentare rilevanti rischi nei confronti dei diritti degli interessati relativamente ai propri dati personali e che, di conseguenza, tali diritti devono essere tutelati attraverso l'applicazione delle regole del GDPR.

Gli articoli più rilevanti del GDPR che trovano applicazione nei sistemi di IA sono i seguenti:

- ➔ Gli articoli 13 e 14 relativi al principio di trasparenza, da cui discende l'obbligo, per il titolare che intende trattare i dati, anche avvalendosi di processi decisionali automatizzati, di informare l'interessato di tale processo, fornendogli altresì informazioni significative sulla logica utilizzata". Tale norma è direttamente collegata al dibattito relativo al tema della *explainability* dell'IA (capitolo 31).
- ➔ L'articolo 22 che sancisce il diritto dell'interessato a non essere oggetto di un processo decisionale automatizzato, compresa la profilazione, salvo che tale processo: :
  - ◆ sia necessario per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento;
  - ◆ sia autorizzato dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento [...];
  - ◆ si basi sul consenso esplicito dell'interessato (base giuridica la cui dimostrazione è molto complessa in quanto il titolare è tenuto a dimostrare che tale consenso era libero e consapevole; inoltre il titolare deve sempre permettere la revoca del consenso, gestendo nel caso gli effetti della revoca stessa sul trattamento dei dati).
- ➔ L'articolo 25 in forza del quale tutti i processi che trattano dati personali devono sempre rispettare i principi della privacy by design e privacy by default: ne deriva che il rispetto dei principi per il trattamento dei dati deve essere tenuto in considerazione già in fase di progettazione e realizzazione del sistema di IA
- ➔ Gli articoli 44 e seguenti che dettano le regole per i trasferimenti dei dati personali al di fuori dello SEE, trattamento molto frequente nei sistemi di IA che operano sul cloud.

Per cogliere le sfide dell'IA rispetto alla protezione dei dati personali esistono molti documenti a livello internazionale e europeo. In questa sede per ragione di spazio si è deciso di citare

solo l'Opinion 4/20 dell'EDPS[133] che evidenzia e analizza alcuni profili problematici del rapporto tra GDPR e IA:

1) opacità dell'IA che limita il principio di trasparenza a favore degli utenti che subiscono le decisioni dell'IA senza essere in grado di coglierne appieno le logiche; tale rischio, a volte, coinvolge anche gli stessi produttori dei sistemi (vedere capitolo 31);
2) limitazioni del campo di applicazione della legislazione UE esistente;
3) cambiamenti del funzionamento dei sistemi di IA;
4) incertezza riguardo l'attribuzione delle responsabilità (vedere capitolo 33).

Degno di nota è altresì l'articolo 12 della proposta di Regolamento "sui principi etici per lo sviluppo, la diffusione e l'utilizzo dell'intelligenza artificiale, della robotica e delle tecnologie correlate" contenuto nella Relazione[134] del Parlamento europeo, nel quale, in riferimento al rapporto tra l'intelligenza artificiale e la vita privata e la protezione dei dati personali, afferma che "*L'uso e la raccolta di dati biometrici a scopo di identificazione a distanza nei luoghi pubblici, come il riconoscimento biometrico o facciale, comportano rischi specifici per i diritti fondamentali e sono diffusi o utilizzati solo dalle autorità pubbliche degli Stati membri per finalità di interesse pubblico rilevante. Dette autorità garantiscono che tale diffusione o utilizzo siano comunicati al pubblico, proporzionati, mirati e limitati a obiettivi e a un luogo specifici nonché limitati nel tempo, in conformità del diritto dell'Unione e nazionale, in particolare del regolamento (UE) 2016/679 e della direttiva 2002/58/CE, e nel debito rispetto della dignità e dell'autonomia umane e dei diritti fondamentali sanciti dalla Carta, segnatamente il diritto al rispetto della vita privata e alla protezione dei dati personali".*

30.2 Privacy e IA a livello internazionale

Un documento molto interessante è la "Guidance on AI and data protection", pubblicata da ICO[135] (autorità privacy inglese) nel 2020.

Tale guida analizza in maniera specifica i seguenti temi:

- la scarsa trasparenza degli algoritmi con cui i sistemi di IA arrivano a prendere decisioni[136];
- la sicurezza;
- la necessità e la proporzionalità dei trattamenti e la minimizzazione dei dati;
- l'esercizio dei diritti da parte dell'interessato.

[133] https://edps.europa.eu/data-protection/our-work/publications/opinions/edps-opinion-european-commissions-white-paper_en.
[134] Proposta di risoluzione recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate (2020/2012(INL), disponibile presso: https://www.europarl.europa.eu/doceo/document/A-9-2020-0186_IT.html#title1.
[135] https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-artificial-intelligence-and-data-protection/.
[136] Vedere anche: https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-ai/.

Riguardo alle “sfide” dell’IA e della protezione dei dati personali, Mutschelknaus[137] sottolinea che nella progettazione di una soluzione di IA occorre tenere in conto alcune peculiarità, ad esempio:

- come e quando i dati raccolti vengono anonimizzati;
- nel caso in cui la base legale sia il consenso, come ottenerlo in modo che sia valido anche considerando che si potrebbe pensare a un nuovo uso dei dati raccolti senza che questo debba comportare una nuova richiesta del consenso e rispettando il principio di limitazione della finalità sugli ulteriori trattamenti per "finalità non incompatibili";
- garantire la possibilità di revoca del consenso, tenendo anche conto che le soluzioni l’evoluzione di sistemi di IA;
- progettare soluzioni di IA che permettano la cancellazione dei dati personali.

30.3 Il caso INPS

In relazione ai profili di protezione dei dati nei sistemi di IA, appare molto interessante un procedimento aperto dal Garante privacy nei confronti dell’INPS per un non corretto utilizzo del sistema denominato Savio: attraverso tale software l’INPS analizzava i certificati medici presentati dai lavoratori pubblici e privati e attribuiva al lavoratore un punteggio circa il “grado di propensione all’assenza per malattia ingiustificata”, creando in questo modo una profilazione del lavoratore.

In relazione all’utilizzo di tale software di IA, il Garante privacy aprì procedimento contro l’INPS (sulla base del D. Lgs. 196/2003, Codice privacy vigente all’epoca) con il quale si contestava di aver creato una profilazione dei dipendenti basata sui loro dati di salute senza aver rispettato gli adempimenti privacy: più esattamente senza aver effettuato la notifica al Garante (allora richiesta dall’art. 37 del Codice), senza aver sottoposto il trattamento alla verifica preliminare ex art. 17 e senza aver informato di tale trattamento i dipendenti. Conseguentemente, con il **provvedimento numero 429 del Garante privacy del 29 novembre 2018**, veniva comminata all’INPS una sanzione amministrativa pecuniaria di 40.000 euro con il correlativo blocco circa l’utilizzo del sistema.

[137] https://insidebigdata.com/2020/07/23/top-five-data-privacy-issues-that-artificial-intelligence-and-machine-learning-startups-need-to-know/.

Intervista a Guido Scorza

Componente del Collegio del Garante per la protezione dei dati personali, avvocato, giornalista pubblicista. Già responsabile degli affari regolamentari del team per la trasformazione digitale della Presidenza del Consiglio dei ministri e, poi, consigliere giuridico del Ministro per l'innovazione. È autore di alcuni libri, gli ultimi dei quali sono "Processi al futuro", 2020, Egea, e "Intelligenza artificiale. L'impatto sulle nostre vite, diritti e libertà" con A. Longo, 2020, Mondadori.

Domanda 1. Che cos'è l'intelligenza artificiale? Perché è una sfida complessa per la società?

Definire l'intelligenza artificiale è una sfida nella sfida. Esiste una moltitudine di definizioni capaci di identificarne una o più caratteristiche.

Probabilmente, in questo contesto, è sufficiente pensare all'intelligenza artificiale come a un approccio tecnologico - più che a specifiche tecnologie - destinate ad avere sull'umanità un impatto superiore a quello avuto da Internet in termini di pervasività e impatto sul modo in cui viviamo, lavoriamo, governiamo mercati e Nazioni. La complessità della sfida è data da elementi diversi: la velocità dell'evoluzione tecnologica è uno dei principali perché rende difficile che uomini, culture e regole riescano ad adeguarsi alla trasformazione che la diffusione delle soluzioni di intelligenza artificiale impone, la pervasività è un altro elemento perché l'impatto dell'intelligenza artificiale investe contemporaneamente ogni ambito della vita dell'uomo.

Domanda 2. **In quali ambiti già adesso è applicata in Italia l'intelligenza artificiale?**

Difficile tracciare un perimetro. La linea di confine è un po' funzione della definizione che si adotta. Gli ambiti di applicazione già attuale sono diversissimi e vanno dal marketing, alla meccanica industriale, passando per l'assistenza alla clientela (attraverso le chatbot), alla medicina con particolare riferimento agli strumenti di diagnosi sino ad arrivare alla pubblica amministrazione che inizia a utilizzare gli algoritmi, sebbene ancora timidamente, per la gestione di procedimenti amministrativi a bassa discrezionalità.

Domanda 3. **Come coniugare la tecnologia IA con i diritti alla protezione dei dati personali?**

Personalmente credo che la strada sia quella indicata dal GDPR attraverso i principi del by design e by default.

Solo se si riuscirà a fare in modo che il rispetto della protezione dei dati sia un vincolo allo sviluppo delle soluzioni di intelligenza artificiale sin dalla progettazione si riuscirà nella sfida.

Credo sia pressoché impossibile ipotizzare di coniugare le potenzialità dell'intelligenza artificiale e il rispetto della privacy in una logica ex post. Quella degli interessati e delle autorità di protezione diverrebbe una rincorsa continua e inutile.

Domanda 4. **Quali consigli e quali percorsi può indicare per i produttori e per le imprese che utilizzano i servizi di IA?**

Considerare il rispetto dei diritti umani parte integrante della sfida tecnologica che abbiamo davanti.

Così come stiamo giustamente investendo ogni nostra risorsa nello sviluppo e utilizzo di soluzioni di IA sempre più efficienti e performanti, dobbiamo pensare al fatto che tali soluzioni rispettino la privacy e gli altri diritti dell'uomo come uno dei tanti elementi capaci di accrescere l'appeal della tecnologia che produciamo o utilizziamo.

Questo per ragioni diverse.

Innanzitutto perché in caso contrario, prima o poi, la tecnologia che abbiamo prodotto o che intenderemmo utilizzare potrebbe scontrarsi con le regole o, forse, peggio, produrci un danno all'immagine finendo travolta in una qualche tempesta mediatica dovuta al suo scarso rispetto dei principi e valori etici come, in alcuni casi, è già successo.

E, poi, perché in un intervallo medio breve, probabilmente, gli utenti e consumatori impareranno ad apprezzare le tecnologie capaci di garantire loro il massimo risultato in termini di semplificazione della vita con la minima possibile compressione dei loro diritti, a cominciare dalla privacy.

L'esperienza dell'attenzione dei consumatori verso scelte ecosostenibili dovrebbe essere d'insegnamento.

Senza dire che, spesso, in materia di intelligenza artificiale e big data, più protezione dei dati personali significa anche maggior qualità dei dati utilizzati dalla tecnologia di riferimento e, di conseguenza, maggiore efficienza.

Domanda 5. Quali iniziative il Garante italiano può mettere in campo in sinergia con gli altri Garanti?

Si tratta di guidare una trasformazione epocale della società senza ostacolarla ma facendo in modo che essa produca benefici quanto più diffusi possibili per l'umanità.

Sfortunatamente è una sfida che non ha precedenti nella storia con la conseguenza che dovremo imparare o provare a imparare giorno dopo giorni dagli errori di approccio che inesorabilmente commetteremo.

In cima alla lista delle cose da fare c'è certamente un investimento mai fatto sin qui in termini di educazione alla cultura della protezione dei dati personali.

Senza, la battaglia è persa in partenza.

Per il resto sarà indispensabile un dialogo costante con le imprese che scongiuri il rischio di autorità che scrivono le regole del futuro chiuse nelle loro torri d'avorio.

Bisognerà impegnarsi, naturalmente, da entrambe le parti.

Domanda 6. **Come assicurare il rispetto del principio di trasparenza nei confronti dei cittadini consumatori e un esercizio effettivo dei diritti ?**

La tecnologia che crea taluni problemi può anche essere la miglior alleata di interessati e autorità per risolverli.

Le interfacce, il legal design, la multimedialità delle nuove tecnologie sono strumenti ai quali guardare in maniera sempre più insistente.

È una partita che si può vincere se riusciremo a fare comprendere a tutti l'enorme valore degli interessi in gioco.

31 Trasparenza dell'IA

È doveroso ricordare che la stesura del seguente capitolo riflette la situazione normativa nazionale e transnazionale in vigore al 31.12.2020 in termini di requisiti di trasparenza dell'intelligenza artificiale a garanzia della tracciabilità dei sistemi e a dimostrazione delle operazioni compiute dall'algoritmo.

31.1 Il principio della trasparenza dell'IA

Come già anticipato al paragrafo 30.1, il trattamento di dati tramite software di IA deve rispettare il principio di trasparenza, e, nello specifico, spiegare la logica del software.

Già nel 2017 l'Information Commissioner's Office (ICO) del Regno Unito si riferiva al ruolo degli algoritmi nel processo decisionale descrivendolo come "responsabilizzazione algoritmica»" e affermando come fosse cruciale "*essere in grado di verificare che gli algoritmi utilizzati e sviluppati dai sistemi di apprendimento automatico stiano effettivamente facendo ciò che noi pensiamo facciano e non producano risultati discriminatori, errati o ingiustificati*"[138].

Relativamente alla proposizione di un principio di trasparenza che consenta agli interessati di controllare l'utilizzo che viene fatto con i propri dati, sia forniti direttamente e osservati che dedotti, interessante è la posizione del Garante europeo della protezione dei dati (GEPD o EDPS), il quale già nel suo Parere 7/2015 affermava che "*le persone devono ricevere informazioni chiare su quali dati sono oggetto di trattamento, compresi i dati osservati o dedotti che le riguardano, e devono ricevere informazioni più precise sull'uso e sullo scopo di utilizzazione di tali dati, compresa la logica utilizzata negli algoritmi per determinare le presunzioni e le ipotesi che le riguardano*"[139].

Le decisioni assunte da un algoritmo di IA rientrano senza dubbio nei "processi decisionali automatizzati" (art. 22 del GDPR e Linee Guida WP 29 - Linee Guida sul processo decisionale automatizzato relativo alle persone fisiche). In relazione a tale tipologia di trattamenti l'art. 13 relativo alla c.d. Informativa nell'elencare i contenuti della stessa al comma 2 punto f) stabilisce che il titolare è tenuto a spiegare all'interessato (soggetti i cui dati vengono trattati) "*l'esistenza di un processo decisionale automatizzato, compresa la profilazione [...] e, almeno in tali casi, informazioni significative sulla logica utilizzata, nonché l'importanza e le conseguenze previste di tale trattamento per l'interessato*".

[138] Information Commissioner's Office, "Big data, artificial intelligence, machine learning and data protection," 2017.
[139] Sintesi esecutiva del parere del Garante europeo della protezione dei dati: "La risposta alle sfide dei megadati: richiesta di trasparenza, controllo da parte degli utilizzatori, protezione dei dati fin dalla progettazione e responsabilità", Gazzetta ufficiale dell'Unione europea, vol. 2016/C 067/05, pagg. 13-15, 2016.

Si tratta della cosiddetta *IA spiegabile (explainable AI*), in forza della quale devono essere fornite le informazioni chiave all'utente perché questi possa comprendere, nella maniera più semplice possibile, la ragione poste alla base della decisione che viene adottata dalla macchina che possono impattare sui dati della persona (c.d. *spiegabilità*). Ciò potrebbe anche permettere all'utente di contribuire al processo decisionale (ad esempio, inviando alla macchina informazioni in grado di correggere eventuali errori di valutazione riscontrati).

Tale profilo ha acquistato molta più rilevanza negli ultimi anni, essendosi incrementato lo sviluppo di sistemi di IA basati sul *deep learning*, una tipologia di apprendimento caratterizzata da reti neurali complesse che rendono spesso inaccessibili e non spiegabili le decisioni adottate dalla macchina: in questi casi, in sostanza, il rischio è che l'utente non sia messo in grado di capire come funziona l'algoritmo che tratta i suoi dati. Infatti, "*lo sviluppo dell'IA ha enfatizzato la qualità in termini di accuratezza e di generalizzazione, invece che di comprensione e validazione*"[140], con il rischio che vengano creati dei modelli il cui funzionamento possa sfuggire anche agli esperti. Proprio al fine risolvere il problema dell'opacità dell'algoritmo (il cosiddetto *problema della scatola nera* o *black-box effect*), di recente sono stati portati avanti studi e lavori di ricerca che cercano di ovviare a questo deficit.

Tra questi lavori rientrano certamente le linee guida della Commissione Europea sull'etica per l'IA[141]. All'interno del documento, il gruppo di lavoro fornisce indicazioni chiare su come realizzare un'IA affidabile elencando sette requisiti che i sistemi di IA dovrebbero soddisfare e per la cui attuazione possono essere utilizzati metodi tecnici e non tecnici (vedere capitolo 29).

Per quanto concerne nello specifico il requisito della "trasparenza", il documento precisa che si compone di tre elementi distinti:

1. tracciabilità,
2. spiegabilità;
3. comunicazione.

Circa la *tracciabilità*, le linee guida precisano che "i dati e i processi che determinano la decisione del sistema di IA, compresi quelli di raccolta ed etichettatura dei dati, come pure gli algoritmi utilizzati, dovrebbero essere documentati secondo i migliori standard per consentire la tracciabilità e aumentare la trasparenza".

In particolare, la documentazione dovrebbe avere ad oggetto le metodologie, le procedure e le tecniche di programmazione e di addestramento utilizzate per costruire, sottoporre a prova e convalidare i sistemi di IA al fine di evitare distorsioni nel trattamento dei dati che potrebbero dar luogo a discriminazioni a danno dell'interessato. All'interno di detta documentazione rientrerebbero, ad esempio, la descrizione dell'algoritmo, l'obiettivo di ottimizzazione perseguito dal modello e le ponderazioni applicate per determinati parametri.

[140] Fosca Giannotti, Dino Pedreschi. "Explainable AI". Aprire le scatole nere per una Intelligenza Artificiale umana. *Gnosis. Rivista italiana di intelligence*. 2019. Vol. 2, pp. 36–45.
[141] Gruppo indipendente di esperti ad alto livello sull'intelligenza artificiale. *Orientamenti etici per un'IA affidabile (Ethics guidelines for trustworthy AI)*. Bruxelles: Commissione europea, 2019. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

Tale documentazione, infatti, consentirebbe di capire la ragione posta alla base di una decisione errata del sistema e, quindi, di prevenire errori futuri. A tal fine, i registri, la documentazione e, se del caso, i set di dati dovrebbero essere conservati per un periodo di tempo ragionevole al fine di garantire l'effettivo svolgimento dei controlli e il rispetto della legislazione pertinente. Contemporaneamente, occorrerebbe adottare misure adeguate volte a garantire che tali elementi siano messi a disposizione su richiesta, in particolare, per essere sottoposti a prove o ispezioni da parte delle autorità competenti e, se necessario, dovrebbero essere adottate idonee misure volte a garantire la tutela delle informazioni riservate, come i segreti commerciali.

La *spiegabilità* attiene alla capacità di spiegare sia i processi tecnici di un sistema di IA che le relative decisioni umane, importante soprattutto se il sistema di IA influisce considerevolmente sulla vita delle persone. Affinché un sistema di IA possa essere tecnicamente spiegabile, gli esseri umani devono poter capire e tenere traccia delle decisioni prese dal sistema stesso. Inoltre, suggeriscono le linee guida, tale spiegazione dovrebbe essere tempestiva e adeguata alle competenze del portatore di interesse in questione a prescindere dal fatto che lo stesso sia un non esperto, un'autorità di regolamentazione o un ricercatore. Dovrebbero poi essere disponibili indicazioni sul grado con cui un sistema di IA influenza e plasma il processo decisionale organizzativo, sulle scelte progettuali del sistema, nonché sulla logica posta alla base della sua distribuzione.

In ultimo, per quanto concerne la *comunicazione*, il documento citato precisa che i sistemi di IA non devono presentarsi agli utenti come esseri umani e gli esseri umani hanno il diritto di essere a conoscenza del fatto che stanno interagendo con un sistema di IA. Oltre a ciò, dovrebbero essere comunicate agli operatori – o agli utenti finali – le capacità e le limitazioni del sistema in maniera consona al caso d'uso in questione (ciò, ad esempio, potrebbe comprendere la comunicazione del livello di precisione del sistema di IA e dei suoi limiti).

In conclusione, è possibile notare come la *spiegabilità* diventi a sua volta fondamentale per garantire la consapevolezza (la cosiddetta *awareness*) degli utenti con riferimento agli eventi elaborati e alle decisioni adottate dalla macchina stessa.

31.2 Il caso del Ministero della pubblica istruzione

I profili sopra descritti relativi alla spiegabilità dell'IA hanno trovato un'importante applicazione pratica nella decisione del **Consiglio di Stato 8 aprile 2019 n. 2270**.

Questo il caso: il Ministero della pubblica istruzione, volendo attuare un piano straordinario di assunzioni a tempo indeterminato di personale docente per le istituzioni scolastiche statali, decise di gestire tali procedure di assunzione attraverso un software. Su tale scelta si aprì un contenzioso nel quale si lamentava che le decisioni "assunte dall'algoritmo" non tenevano conto di alcune variabili espressamente previste dalla legislazione (ad esempio la vicinanza della sede all'abitazione della persona) e che il provvedimento di assegnazione era carente di motivazione (ex art. 3 legge 241/1990).

La questione fu decisa con l'importante sentenza sopra citata**.** In tale decisione il giudice amministrativo affermò, in primo luogo, che l'utilizzo nella PA di sistemi automatizzati è del tutto legittima e che in molti casi può contribuire al buon andamento della stessa pubblica amministrazione proprio in virtù dell'assenza dell'intervento umano in grado di alterare il corretto svolgimento dell'attività pubblica.

Dopo aver sancito tale principio, il Consiglio sancì però che una decisione assunta tramite "algoritmo" non è in ogni caso sottratta al rispetto ed all'applicazione dei principi previsti dalla legge n. 241/1990 sul procedimento amministrativo e che tale tipologia di provvedimento deve (come tutti gli altri) poter essere valutato sotto il profilo della corretta motivazione, della ragionevolezza, della pubblicità e della trasparenza. Conseguentemente la pubblica amministrazione è tenuta a svolgere una costante opera di perfezionamento dell'algoritmo mediante test e aggiornamenti periodici, nonché ad assicurare la trasparenza del funzionamento dell'algoritmo stesso.

Per il Consiglio di Stato, la "caratterizzazione multidisciplinare" dell'algoritmo (costruzione che certo non richiede solo competenze giuridiche, ma tecniche, informatiche, statistiche, amministrative) non esime dalla necessità che la "formula tecnica", che di fatto rappresenta l'algoritmo, sia corredata da spiegazioni che la traducano nella "regola giuridica" ad essa sottesa e che la rendano leggibile e comprensibile.

La trasparenza, quindi, deve essere garantita in tutti gli aspetti: dai creatori dell'algoritmo al procedimento usato per la sua elaborazione, al meccanismo di decisione, comprensivo delle priorità assegnate nella procedura valutativa e decisionale e dei dati selezionati come rilevanti. Ciò al fine di poter verificare che i criteri, i presupposti e gli esiti del procedimento automatizzato siano conformi alle prescrizioni e alle finalità stabilite dalla legge o dalla stessa amministrazione e affinché siano chiare – e conseguentemente sindacabili – le modalità e le regole in base alle quali esso è stato impostato.

Tali principi giuridici sono stati poi ripresi e ribaditi in successive decisioni sempre del Consiglio di Stato. (CdS n. 2936/2019; CdS n. 8474/2019 e CdS n. 881/2020).

32 Qualificazione giuridica del prodotto IA

Come già sopra accennato, allo stato attuale non esiste, né a livello comunitario né a livello nazionale, una disciplina legislativa specifica emanata per i software di intelligenza artificiale: ne deriva che ai fini della qualificazione giuridica del software di IA occorrerà tenere conto del quadro legislativo generale.

Sul punto si evidenzia che, in linea generale, i prodotti sono sottoposti alla Direttiva 2001/95/CE sulla sicurezza generale dei prodotti (recepita in Italia nel Codice del consumo), salvo il caso in cui il prodotto, per il suo funzionamento e per la sua destinazione d'uso, non rientri sotto una disciplina specifica, acquisendo quindi specifica qualificazione giuridica.

Ciò vale anche per i software di IA che dovranno quindi rispettare la Direttiva 2001/95/CE, salvo che la loro destinazione d'uso specifica non comporti l'applicazione di una direttiva verticale riferita alla singola tipologia di prodotti. È il caso - solo a titolo di esempio - del software IA contenuto all'interno di un dispositivo medico che, quindi, deve seguire la disciplina verticale dei dispositivi medici (vedere capitolo 22).

33 Responsabilità civile

I profili di responsabilità civile connessi all'impiego dell'intelligenza artificiale sono certamente di grande impatto e rilevanza per lo sviluppo della stessa. Uno dei punti cardine è quello della particolare difficoltà ad ottenere un risarcimento per le vittime, difficoltà connessa al fatto che l'attuale regolamentazione sulla responsabilità civile non pare adattarsi adeguatamente alle caratteristiche peculiari delle nuove tecnologie.

Sul tema si è espresso anche il Parlamento Europeo con una recentissima Risoluzione recante raccomandazioni relativamente al regime di responsabilità civile per l'intelligenza artificiale[142], nella quale è stato affermato come "*procedure eque in materia di risarcimento implicano che ogni persona che subisca un danno cagionato da sistemi di IA o il cui patrimonio sia danneggiato da sistemi di IA benefici dello stesso livello di protezione previsto per i casi in cui non sia coinvolto un sistema di IA*". Di conseguenza l'utente "*deve avere la certezza che per il potenziale danno arrecato dai sistemi che utilizzano l'IA esistano un'adeguata copertura assicurativa e una via legale definita per il risarcimento*".

[142] Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione su un regime di responsabilità civile per l'intelligenza artificiale (2020/2014(INL)), disponibile presso https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_IT.pdf.

La difficoltà di valutare la responsabilità civile della AI con gli attuali strumenti giuridici era già stata rilevata dalla Commissione Europea[143] che aveva palesemente affermato che "*La combinazione di componenti digitali diversi in un ecosistema complesso e la pluralità di soggetti coinvolti possono rendere difficile accertare l'origine di un potenziale danno e il soggetto responsabile dello stesso*".

L'AI presenta infatti peculiarità specifiche: da un lato la capacità della macchina di sviluppare diversi livelli di autonomia, attraverso una costante evoluzione dei processi di autoapprendimento (e di conseguenza decisionali) che non si arrestano dopo l'uscita dalla sfera di controllo del produttore; dall'altro, la complessità dei sistemi algoritmici su cui si esprime la prestazione dell'IA e che rende taluni risultati non prevedibili ex ante e, talvolta, neppure spiegabili o comprensibili ex post.

Sia sul piano etico che giuridico, con riguardo ai criteri di riparto delle responsabilità, la situazione si complica ulteriormente sul fronte della **imputabilità soggettiva**, laddove si prendano in considerazione sia la pluralità di interventi nello sviluppo di prodotti e servizi IA, sia le prospettive dettate dall'Internet of Things, dove entrano in gioco innumerevoli attori (p.e. i produttori dei singoli dispositivi connessi, i programmatori dei software e i gestori della rete sulla quale i terminali comunicano).

In buona sostanza, le caratteristiche dell'IA non appaiono più coerenti con gli usuali criteri di allocazione della responsabilità tra produttori, così come tra questi e gli utilizzatori, con la conseguenza che l'applicazione delle norme tradizionali in tema di responsabilità civile lascia aperte numerose questioni fondamentali.

Tra queste senz'altro rilevano: la prevedibilità di modifiche del prodotto (p.e. difetto successivo), l'incidenza di azioni o omissioni dell'utilizzatore (concorso di colpa), l'intelligibilità del funzionamento dell'IA (anche con riferimento all'autoapprendimento) e le conseguenze sulla prova della colpa (specie per le macchine nelle quali vengono integrati algoritmi di deep learning, che le dotano di capacità decisionali fondate sull'esperienza e sull'auto-apprendimento).

La scarsa adeguatezza dell'attuale quadro giuridico a garantire il risarcimento danni potrebbe poi:

- disincentivare i consumatori dall'utilizzo di beni o servizi IA, a fronte della permanenza di alti livelli di sfiducia connessi all'ancora indefinita tutela risarcitoria oltre che al timore per l'abuso dei dati personali raccolti e a una percezione di alto rischio di manipolazione nelle scelte di acquisto[144];
- scoraggiare gli investimenti, frenando lo sviluppo nel settore.

Appare pertanto necessario e urgente procedere a riformare l'impianto normativo vigente – ad ogni livello: internazionale, europeo o nazionale – per adattarlo alle nuove tecnologie, oltre che uniformare le legislazioni nazionali in materia, in modo da favorire gli scambi transfrontalieri UE, scongiurando la frammentazione regolamentare tra gli Stati membri.

Un importante passo in questa direzione è stato di recente effettuato dal Parlamento Europeo tramite la già menzionata Risoluzione relativa proprio sul regime di responsabilità civile per

[143] *Relazione sulle implicazioni dell'intelligenza artificiale, dell'Internet delle cose e della robotica in materia di sicurezza e di responsabilità*. Bruxelles: Commissione europea, 2020. https://ec.europa.eu/info/publications/commission-report-safety-and-liability-implications-ai-internet-things-and-robotics_en

[144] https://www.beuc.eu/publications/beuc-x-2020-078_artificial_intelligence_what_consumers_say_report.pdf

l'IA: la risoluzione presenta poi in allegato una proposta di Regolamento, che una volta approvato troverà diretta applicazione nei paesi membri dell'Unione.

La proposta di Regolamento presenta poi una architettura giuridica basata su un approccio denominato RMA (*risk management approach*): in particolare nel testo di regolamento proposto viene delineato un sistema di responsabilità puramente oggettiva per gli operatori di sistemi di IA ad alto rischio (sancendo per questi l'obbligo di stipulare un'assicurazione per la responsabilità civile idonea a coprire danni fino ad un importo massimo di 2 milioni di euro nel caso di morte o danni alla salute o all'integrità fisica e fino ad 1 milione di euro nel caso di danni al patrimonio) ed un sistema di responsabilità per colpa per gli operatori di sistemi di IA che non rientrano nella nozione di alto rischio.

E' definito poi 'alto rischio': "*un potenziale significativo in un sistema di IA che opera in modo autonomo di causare danni o pregiudizi a una o più persone in modo casuale e che va oltre quanto ci si possa ragionevolmente aspettare*". A titolo di esempio potremmo qualificare come ad alto rischio l'utilizzo di un police robot capace di fermare persone e eseguire arresti in una piazza pubblica, perché da un lato coinvolge un numero di persone potenzialmente ignota e non verificabile e dall'altro ha un impatto su diritti fondamentali quali la libertà e l'integrità personale.

La Risoluzione sopra indicata, riprendendo poi alcuni concetti di un interessante studio del Parlamento europeo su IA e responsabilità civile[145] del luglio 2020, stabilisce che le riforme dirette ad adattare l'impianto normativo vigente dovrebbero far ricadere la responsabilità sul soggetto che è maggiormente in grado di

i. identificare il rischio;
ii. controllare e minimizzare il rischio
iii. gestire lo stesso.

Circa poi i soggetti che potranno essere chiamati a rispondere, la Risoluzione distingue tra:

- "operatore di front-end": la persona fisica o giuridica che esercita un certo grado di controllo su un rischio connesso all'operatività e al funzionamento del sistema di IA e che beneficia del suo funzionamento;
- "operatore di back-end": la persona fisica o giuridica che, su base continuativa, definisce le caratteristiche della tecnologia e fornisce i dati e il servizio di supporto di back-end essenziale e pertanto esercita anche un elevato grado di controllo su un rischio connesso all'operatività e al funzionamento del sistema di IA.

Alla luce di quanto sopra, quindi, si ritiene che strumenti essenziali per la corretta allocazione della responsabilità per danno saranno:

i. l'obbligo di analisi del rischio a carico del produttore – sia preventiva all'immissione sul mercato sia successiva a eventuali modifiche rilevanti dell'applicazione di IA durante il relativo ciclo di vita – che tenga conto dell'uso previsto, dell'uso prevedibile e, se del caso, dell'uso scorretto ragionevolmente prevedibile del prodotto, e che sia anche estesa al caso di interazione con altri dispositivi;
ii. l'obbligo di cooperazione con il produttore nell'individuazione del rischio a carico di tutti gli operatori economici che partecipano alla specifica catena di valore;
iii. la previsione di requisiti di trasparenza degli algoritmi;
iv. obblighi specifici di sorveglianza umana, sin dalla progettazione e per tutto il ciclo di vita dei prodotti e dei sistemi di intelligenza artificiale.

[145] https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU(2020)621926.

34 IA e diritto d'autore

Le IA sono costituite da una struttura “fisica” e da una struttura “intellettuale”. La parte fisica sono i meccanismi che la compongono (che possono essere antropomorfi, zoomorfi, amorfi o un ibrido di tali forme). E’ considerata dotata di una parte “fisica” anche un’IA interamente su cloud poiché occupa una o più partizioni di memoria che costituiscono un substrato fisico. La parte intellettuale è invece costituita dai programmi informatici che ne consentono il funzionamento.

Questa distinzione è fondamentale quando si affronta il tema dei diritti di proprietà intellettuale legati alla progettazione e allo sviluppo dell’IA.

Il software, e quindi gli algoritmi, che stanno alla base del funzionamento delle tecnologie intelligenti sono, infatti, tipicamente protetti ai sensi della legge sul diritto d’autore, che annovera i programmi per elaboratore tra le opere dell’ingegno di carattere creativo, tutelati alla stregua delle opere letterarie, “in qualsiasi forma espressi, purché originali quale risultato della creazione intellettuale dell’autore” (Legge 633/1941, art. 2, comma 8).

Come per le altre opere dell’ingegno di carattere creativo, la legge ricollega la nascita dei diritti d’autore alla creazione della stessa opera, senza necessità di alcuna formalità[146].

Tali diritti sono riconducibili a due grandi macro-categorie: i diritti di natura morale (vale a dire, il diritto alla paternità intellettuale dell’opera, il diritto alla sua integrità, il diritto di inedito e il diritto al ritiro dell’opera dal commercio) e i diritti di natura patrimoniale (vale a dire, i diritti di sfruttamento economico dell’opera, che, per il software, sono declinati, dall’art. 64-bis della legge sul diritto d’autore, nel diritto di riproduzione, nei diritti di traduzione, adattamento, trasformazione e ogni altra modificazione del programma, e, infine, nel diritto di distribuzione).

La differenza fondamentale tra diritti morali e diritti di sfruttamento economico risiede nel fatto che i primi sono diritti di carattere personale, perpetui, inalienabili, imprescrittibili ed irrinunciabili, laddove, i secondi hanno durata limitata nel tempo e sono diritti disponibili: possono, cioè, costituire oggetto di cessione, rinuncia e alienazione.

[146] Si segnala, per approfondimenti la seguente pubblicazione. Joint Research Centre (JRC). *Intellectual property and artificial intelligence - A literature review*. Luxembourg: Publications Office of the European Union, 2019.

34.1 Brevettare un'IA

Se lo sviluppatore di un algoritmo intelligente gode, ex lege ed automaticamente, della tutela garantita dalla normativa sul diritto d'autore, che gli attribuisce, fin dalla creazione del software, i diritti morali e patrimoniali di cui si è detto, non altrettanto accade con la tutela brevettuale.

Secondo il Codice della proprietà industriale (D. Lgs. 30/2005), il software in quanto tale e i metodi matematici non possono essere considerati alla stregua di invenzioni e, quindi, non sono brevettabili nella misura in cui la domanda di brevetto o il brevetto concerne metodi, programmi e presentazioni di informazioni considerati in quanto tali (art. 45).

Il solo software, in sé considerato, non può, dunque, godere della tutela brevettuale. È brevettabile, invece, un algoritmo che sia inserito nel contesto di una invenzione e che presenti i cinque requisiti previsti dalla normativa per la concessione del brevetto: la liceità, l'originalità, la materialità, la novità, l'industrialità.

L'Ufficio europeo dei brevetti (EPO – European patent office) ha integrato le sue linee guida per l'esame delle domande[147] con riferimento alle IA, confermando che sono brevettabili invenzioni riferite ad IA solo se:

- abbiano un carattere tecnico, cioè caratteristiche che contribuiscono alla soluzione di un problema tecnico specifico;
- soddisfino i requisiti di chiarezza, concisione e siano connotate da una descrizione sufficiente, a un qualsiasi esperto del settore, per poter attuare l'invenzione, leggendo la domanda di brevetto.

34.2 IA titolare di diritti di autore e di brevetti

34.2.1 IA titolare di diritti di autore

Alcune applicazioni dell'IA sono in grado di sviluppare del tutto autonomamente opere creative di tipo artistico, letterario o musicale e, conseguentemente, si pongono problemi inerenti all'attribuzione dei diritti riconosciuti dalla normativa in materia di diritto d'autore. In buona sostanza: a chi competono i diritti morali e di sfruttamento economico di cui si è detto, nell'ipotesi in cui l'opera sia stata creata dall'AI in piena autonomia?

Così come per altri ambiti legati allo sviluppo e all'utilizzo di tecnologie IA, la legislazione esistente evidenzia delle lacune che non permettono di inquadrare agevolmente questa complessa tematica.

[147] https://www.epo.org/law-practice/legal-texts/html/guidelines2018/e/g.htm.

Al momento, la possibilità di considerare l'IA creatrice dell'opera come "autore" ai sensi della normativa, riconoscendole, quindi, ogni conseguente diritto, appare preclusa in virtù di due elementi. Il primo è la riconoscibilità dello status di autore generalmente riferito dalle normative – tanto quelle dei Paesi di common law, quando quelle dei paesi di civil law – esclusivamente a una persona fisica; il secondo è il requisito della creatività dell'opera quale condizione per l'accesso alla tutela.

Il Copyright Office statunitense, per esempio, nelle sue linee guida, esclude espressamente che si possa procedere alla registrazione di opere prodotte da animali, piante, macchine o meri processi meccanici che operano automaticamente senza nessun impulso creativo o intervento di un autore umano (Compendium of U.S. Copyright Office Practices).

Sotto questo profilo, non c'è dubbio che l'IA, in quanto "oggetto" e non "soggetto" umano, pur essendo autrice dell'opera, non possa essere diretta destinataria di diritti patrimoniali e meno ancora di diritti morali.

Inoltre, qualunque opera dell'ingegno, ai fini della tutela, deve essere contraddistinta da una originalità – seppur minima (come dal *dictum* della Corte Suprema Americana, nel caso *Feist Publications, Inc., v. Rural Telephone Service Co.*, 499 U.S. 340, 1991) – che scaturisce dalla creatività del suo autore. Se l'autore può essere solo una persona fisica, ne consegue che un'IA non potrà mai essere creativa (Corte giustizia UE, 1 marzo 2012, causa C-604/10, *Football Dataco v. Yahoo! UK*; 16 luglio 2009, causa C-5/08, *Infopaq International A/S v. Danske Dagblades Forening*).

Tuttavia – fintanto che non si scelga una strada sul riconoscimento di una qualche "tipologia" di personalità a una IA totalmente autonoma, prospettiva che, allo stato, dovrebbe escludersi, anche sulla base delle indicazioni da ultimo provenienti dalle istituzioni europee[148] – si potrebbe prospettare il riconoscimento, in via legislativa, di uno specifico diritto 'sui generis' in favore dei progettisti dell'IA per i risultati creativi generati autonomamente dall'IA, come per i software generati con l'aiuto di computer (vedere la *Section* 178 del UK Copyright Designs and Patents Act). Una protezione di questo tipo – analoga a quella del costitutore delle banche dati prevista in Italia dall'art. 102-bis della Legge 633/1941 – risponderebbe al precipuo obiettivo di consentire una remunerazione degli investimenti che sono stati necessari per lo sviluppo di un tale tipo di creatività che, seppur non di matrice umana, nondimeno possono avere notevole ritorno economico.

Al di fuori dell'Europa, è interessante la sentenza del Tribunale della città cinese di Shenzhen, che ha riconosciuto la tutela garantita dal diritto d'autore a un articolo scritto da un programma di intelligenza artificiale. L'articolo era stato creato dal programma di scrittura Dreamwriter, prodotto dal gigante tecnologico cinese Tencent, ma pubblicato da un'altra società che sosteneva la tesi tradizionale secondo cui l'articolo riprodotto non avrebbe potuto considerarsi protetto dal diritto d'autore in quanto non scritto da un essere umano, con la conseguenza che esso avrebbe dovuto considerarsi di dominio pubblico e utilizzabile da chiunque. Nonostante tale tesi difensiva, la Corte, ritenendo soddisfatti i requisiti per beneficiare della tutela del

[148] Relazione sui diritti di proprietà intellettuale per lo sviluppo di tecnologie di intelligenza artificiale del Parlamento europeo del 2 ottobre 2020.

copyright, ha statuito che[149]: “La forma di espressione di cui all’articolo è conforme ai requisiti dell’opera scritta e contiene l’analisi, la selezione e una valutazione sui dati e informazioni rilevanti. Inoltre la struttura dell’articolo era ragionevole, la logica chiara e lo stesso era dotato di una certa originalità”.

34.2.2 IA titolare di brevetti

Un’IA non può essere titolare di un brevetto, né essere riconosciuta autrice del medesimo.

Tanto si desume dalla decisione dell’Ufficio brevetti e marchi degli Stati Uniti (USPTO) rispetto alla richiesta, avanzata dal proprietario dell’IA autonoma DABUS (device for the autonomous bootstrapping of unified sentience), di brevettare due invenzioni. Nel primo caso, la richiesta riguardava il riconoscimento della qualità di inventore e titolare del brevetto in capo all’IA. Nel secondo caso, la richiesta riguardava il solo riconoscimento del diritto morale all’IA, mentre la titolarità del brevetto sarebbe stata del proprietario dell’IA, alla stessa stregua del riconoscimento al datore di lavoro delle invenzioni del dipendente (previsto anche in Italia dall’articolo 2590 del Codice civile italiano).

L’USPTO ha negato il riconoscimento di qualsiasi diritto in capo all’IA, rilevando come la normativa di riferimento attribuisca tutela alla sola persona fisica.

Anche l’EPO ha respinto le medesime due domande presentate anche per il brevetto europeo. Questo sulla base della constatazione che l’inventore debba essere necessariamente un essere umano (art. 81 e punto 19 dell’EPC - *European Patent Convention*). Neppure una persona giuridica può essere riconosciuta come inventore, possibilità affrontata e respinta, a suo tempo, nella redazione dell’EPC.

I sistemi di IA, seppur dotati di un identificativo (che non è un nome nel senso giuridico proprio), sono al momento privi di personalità giuridica e quindi non possono assumere diritti ed obblighi. Non possono, conseguentemente, essere configurati quali dipendenti per cui non sono giuridicamente in grado di trasferire il diritto al brevetto (sia volontariamente, sia quando ciò sia previsto dalla normativa vigente).

Infruttuoso, per le stesse motivazioni, un ulteriore tentativo dell’inventore di DABUS presso l’Ufficio della proprietà intellettuale del Regno Unito (UKIPO).

Quanto detto attiene ai profili formali della domanda di brevetto ma non entra nel merito della brevettabilità di quanto richiesto. In definitiva l’invenzione che produce un innalzamento della tecnica è ancora un argomento da definire compiutamente, riservando all’IA il ruolo di strumento – anche molto sofisticato – che risponde alla logica impostata dal progettista degli algoritmi posti alla base della sua intelligenza.

[149] Nanshan District People’s Court, Shenzhen, Guangdong Province, (2019) Yue 0305 Min Chu No. 14010 Civil Judgment. November 24, 2019.

Nell'ipotesi in cui si possa provare che l'IA è libera di spaziare oltre confini delimitati dalla sua programmazione, andrebbe scrutinata una volontarietà di tale comportamento, tornando in campo i concetti – anche filosofici – di creatività, non legata a un'eccezionale capacità combinatoria, quanto a un processo immaginativo, a un'intuizione che non ha nulla a che vedere con il riconoscimento di una qualche forma di soggettività giuridica all'IA.

Considerati i consistenti investimenti in materia – e l'enorme mercato potenziale – lasciando da parte ogni ragionamento di assimilazione in tutto o in parte ad un essere umano (ed a quanto lo caratterizza), una soluzione percorribile appare quella di modificare la disciplina, introducendo una serie di previsioni specifiche per i brevetti generati da un sistema IA.

https://www.youtube.com/watch?v=quxyK5vW_lk&list=PLkNEbE0KewNefnoZZkK7LlcgZJ0oPdO-C

"Copyright in the Age of A.I."

Una intera playlist di video in lingua inglese, prodotti dal "US Copyright Office" imperniata sulla questione del diritto d'autore per opere create da IA.

(durata complessiva: più di 6 ore)

34.3 Il caso Cineca

Un caso che merita di essere analizzato è quello noto come caso "Cineca" che ha portato alla sentenza del **Consiglio di Stato 2 gennaio 2020, n. 30**.

La questione è sorta tra il MIUR, un gruppo di concorrenti e il Cineca – Consorzio interuniversitario.

In questa vicenda, i partecipanti al concorso contestarono che la valutazione della prova scritta, avvenuta attraverso un apposito software (realizzato dal Cineca e presente sul sito del Ministero), generò molti esiti negativi perché il software non rispettava quanto previsto nella disciplina della prova stessa.

I ricorrenti chiesero dunque di poter accedere ai codici sorgenti del software al fine di "decodificare" il meccanismo decisionale dell'algoritmo.

In primo grado la sentenza del Tar Lazio Roma sez. III bis 6 giugno 2019 n. 7333 accolse la richiesta di accesso dei ricorrenti, negando al Cineca la partecipazione al procedimento giudiziario. La sentenza venne impugnata davanti al Consiglio di Stato che, al contrario, riconobbe al Cineca la qualità di controinteressato.

La causa, rinviata al TAR e ancora oggi pendente, appare interessante perché introduce il tema (non ancora risolto) della necessità di trovare un punto di equilibrio tra il principio della trasparenza del provvedimento amministrativo (vedere capitolo 27) e la necessità di tutelare chi progetta software in forza del diritto d'autore.

35 Etica dell'IA

Qualunque sia il modello di impiego dell'IA, la maggior parte degli interpreti punta l'attenzione sul tema etico dell'IA, in particolare in merito alle scelte che questa può esercitare autonomamente.

Alcune decisioni assunte autonomamente dall'IA potrebbero infatti essere valutate - secondo criteri umani - profondamente negative e per alcuni potrebbero giungere a rappresentare un pericolo per la stessa sopravvivenza del genere umano. È infatti innegabile che una decisione assolutamente razionale non è automaticamente "giusta" (si possono qui citare gli esempi di scuola: le auto a guida automatica quando devono scegliere se sacrificare il conducente o i pedoni e le armi robotizzate intelligenti rispetto agli attacchi suicidi).

Non vi è dubbio che lo sviluppo della IA possa comportare il rischio di disfunzioni quali:

1) i pregiudizi dell'algoritmo (vedere capitolo 38);
2) l'emersione di risultati discriminatori (vedere capitolo 38);
3) la mancanza di trasparenza del sistema decisionale (vedere capitolo 31);
4) l'impiego massivo di dati personali che, dalla profilazione, finiscono per condizionare le scelte degli interessati stessi[150] (come dimostrato dal caso di Cambridge Analytica trattato nel paragrafo 39.2).

Rispetto a questi profili, l'Unione europea sta manifestando un impegno e un'attenzione particolari espressi anche in precise scelte legislative già assunte o in via di assunzione (vedi Cap. 29). Si pensi solo, a titolo di esempio, al Reg. UE 2016/679 (GDPR) e al relativo diritto dell'interessato di non essere sottoposto a una decisione basata unicamente su un trattamento automatizzato (art. 22), nonché al diritto di essere messo a conoscenza della "logica" in base alla quale funzionano gli algoritmi (art. 13 comma 2 lett. f).

In relazione ai profili più strettamente etici, l'Unione europea ha scelto di appoggiare lo sviluppo di **sistemi di IA di natura antropocentrica.** In altre parole, la Comunità europea ha dichiarato apertamente che sistemi di IA dovranno comunque vedere sempre una persona al centro.

[150] https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/8053813

A tale scopo la Commissione ha istituito un gruppo di esperti ad alto livello sull'IA[151], istituendo contemporaneamente l'Alleanza europea per l'IA[152], una piattaforma multilaterale aperta con oltre 2.700 membri, per fornire un contributo più vasto ai lavori di tale gruppo di esperti.

Nel dicembre 2018 il gruppo di esperti ha pubblicato un primo progetto di orientamenti etici. Sulla base di tale documento, la Commissione UE, in data 8 aprile 2019, ha pubblicato, con la Comunicazione COM(2019) 168[153], i già citati *Orientamenti etici per un'IA affidabile (Ethics guidelines for trustworthy AI)*, in cui si elencano gli orientamenti che dovranno essere seguiti da sviluppatori e fornitori per garantire una IA affidabile per gli utenti.

Gli orientamenti partono dal presupposto che per ottenere una "intelligenza artificiale affidabile" l'IA dovrebbe:

1) rispettare la legge;
2) osservare i principi etici;
3) dimostrare robustezza.

Sulla base di questi tre elementi e dei valori europei, gli orientamenti individuano sette requisiti fondamentali che le applicazioni di IA dovrebbero soddisfare.

- **Intervento e sorveglianza umana.** I sistemi di IA dovrebbero aiutare le persone a compiere scelte migliori e più consapevoli nel perseguimento dei loro obiettivi; la sorveglianza dell'uomo contribuisce poi a garantire che i sistemi di IA non mettano in pericolo l'autonomia umana o provochino altri effetti negativi.
- **Robustezza tecnica e sicurezza.** Perché l'IA sia affidabile è indispensabile che gli algoritmi siano sicuri, affidabili e sufficientemente robusti da far fronte a errori o incongruenze durante tutte le fasi del ciclo di vita del sistema di IA, oltre che adeguatamente capaci di gestire risultati sbagliati.
- **Riservatezza e governo dei dati.** La tutela della riservatezza e la protezione dei dati devono essere garantite in tutte le fasi del ciclo di vita del sistema di IA. Inoltre i sistemi di IA devono essere di qualità elevata. La qualità dei dati utilizzati per addestrare un'IA è fondamentale per le prestazioni dei sistemi di IA: quando si raccolgono dati, questi possono riflettere condizionamenti di tipo sociale o contenere inesattezze, errori e vizi materiali. Questo aspetto deve essere risolto prima di utilizzare un qualsiasi insieme di dati per addestrare un sistema di IA. Deve essere inoltre garantita l'integrità dei dati. I processi e gli insiemi di dati utilizzati devono essere testati e documentati in ogni fase (tra cui la pianificazione, l'addestramento, i test e la distribuzione).
- **Trasparenza.** La tracciabilità dei sistemi di IA dovrebbe essere garantita: è importante registrare e documentare sia le decisioni adottate dai sistemi sia l'intero processo che ha prodotto le decisioni. Questo comprende una descrizione della raccolta e dell'etichettatura dei dati e una descrizione dell'algoritmo utilizzato. In questo contesto dovrebbe essere prevista, per quanto possibile, la spiegabilità del processo decisionale degli algoritmi, adattata alle persone coinvolte.

[151] https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence, già segnalato in precedenza.
[152] https://ec.europa.eu/digital-single-market/en/european-ai-alliance.
[153] https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:52019DC0168.

- **Diversità, non discriminazione ed equità.** I dati utilizzati dai sistemi di IA (sia per l'addestramento sia per il funzionamento) possono essere inficiati da condizionamenti storici involontari, incompletezza e modelli di governance inadatti. Tale rischio può coinvolgere anche il modo in cui sono sviluppati i sistemi di IA: per superare tali problemi può essere utile la creazione di gruppi di progettazione diversificati e l'istituzione di meccanismi per garantire la partecipazione, in particolare dei cittadini, allo sviluppo dell'IA.
- **Benessere sociale e ambientale.** Per ottenere un'IA affidabile si dovrebbe tenere conto del suo impatto sull'ambiente: tutti gli esseri umani, comprese le generazioni future, devono poter beneficiare della biodiversità e di un ambiente abitabile. L'impatto dei sistemi di IA dovrebbe inoltre essere considerato non solo da una prospettiva individuale, ma anche dal punto di vista della società nel suo complesso: andrebbe quindi prestata particolare attenzione all'uso dei sistemi di IA soprattutto nelle situazioni che riguardano il processo democratico, compresi la formazione di opinioni, i processi decisionali politici o i contesti elettorali.
- **Accountability.** Dovrebbero essere previsti meccanismi che garantiscano la responsabilità e l'accountability dei sistemi di IA e dei loro risultati, sia prima che dopo la loro attuazione.

I suddetti requisiti dovrebbero essere applicati a tutti i sistemi di IA nei diversi contesti e settori, tenendo però conto del contesto specifico in cui si applicano e adottando un approccio basato sul rischio (RMA, già introdotto nel capitolo 33) o sull'impatto (p.e. un'applicazione di IA che suggerisce di leggere un libro non adatto comporta molti meno rischi rispetto a una che sbaglia una diagnosi di tumore e potrebbe quindi essere sottoposta a una vigilanza meno rigorosa).

Infine, i prossimi passi dell'Unione europea in questo settore dovranno definire:

1) una fase pilota circa l'applicazione dei requisiti di cui sopra che coinvolgerà i portatori di interessi che sviluppano o utilizzano l'IA, comprese le amministrazioni pubbliche;
2) una consultazione continua dei portatori di interessi e un processo di sensibilizzazione in tutti gli Stati membri e per i diversi gruppi di portatori di interessi, compresi i settori dell'industria e dei servizi.

Anche la già citata Relazione del Parlamento europeo con relativa proposta di Regolamento[154] (vedi Cap 29) si è espressa sulla necessità di rispettare nella creazione dell'IA alcuni principi di natura etica tra i quali ricordiamo:

- valutazione obbligatoria di conformità ai principi di sicurezza;
- trasparenza e responsabilità;
- assenza di distorsioni e discriminazioni;
- parità di genere;
- sostenibilità ambientale;
- rispetto della vita privata e protezione dei dati personali, nonché obbligo di un monitoraggio successivo da effettuarsi tramite autorità nazionali indipendenti di controllo;
- rilascio di certificati europei di conformità etica.

[154] Proposta di risoluzione recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate (2020/2012(INL), disponibile presso: https://www.europarl.europa.eu/doceo/document/A-9-2020-0186_IT.html#title1 .

Si segnala infine che per quanto attiene alla violazione del principio di discriminazione, è già intervenuta una recente decisione del Tribunale di Bologna. Con un'ordinanza della sezione del Lavoro del 31 dicembre 2020, il giudice ha accertato che la società Deliveroo ha posto in essere una condotta discriminatoria, attraverso l'utilizzo di un algoritmo che poneva limiti alle condizioni di accesso alle sessioni di lavoro dei rider.

Più esattamente, l'algoritmo favoriva l'assegnazione dei turni di lavoro ai lavoratori con meno assenze, sulla base di una profilazione degli stessi, e non teneva conto delle diverse motivazioni dell'assenza.
Il tribunale ha condannato Deliveroo - che sta valutando l'appello e che aveva già interrotto l'uso dell'algoritmo - ad un risarcimento di euro 50.000 e alla pubblicazione del provvedimento sul proprio sito internet ed all'interno della sezione FAQ.

Intervista a Alessandro Longo, giornalista e autore

Alessandro Longo è Direttore responsabile di Agenda Digitale, testata giornalistica online, e coautore del libro "Intelligenza artificiale. L'impatto sulle nostre vite, diritti e libertà".

Domanda 1. Tutti parlano ormai di IA, chi evidenziandone i successi e chi mettendo in guardia rispetto a potenziali pericoli. Cosa ne pensi? Cosa possiamo fare per arrivare a prendere decisioni più consapevoli?

L'intelligenza artificiale può essere una forza per il bene: per la ripresa di produttività dell'Occidente, per la riduzione delle diseguaglianze e per aprire a un futuro dove il lavoro sia fonte di realizzazione per tutti. Allo stesso tempo può essere fattore che peggiora tutti i parametri socio-economici: diseguaglianze, discriminazioni, alienazione degli individui. La Commissione europea e i molti sociologi, filosofi, giuristi che stanno analizzando il fenomeno condividono l'idea che l'IA sia una forza spartiacque che può fare molto bene o molto male all'umanità; e che il primo o il secondo esito dipenderà dalla capacità di gestire il cambiamento da parte delle collettività, dei sistemi e degli individui.

L'esito più auspicabile è raggiungibile solo puntando su una diffusa AI Literacy (istruzione/educazione all'artificial intelligence). Il Governo finlandese ne è stato antesignano, con corsi alla popolazione, imitato in questo dal Governo italiano a fine 2020. Il fine è duplice: stimolare l'innovazione e la crescita economica con una proficua adozione dell'IA, in Paesi e aziende, e al tempo stesso dirigerla eticamente per gli interessi collettivi. Obiettivi che vanno assieme, perché l'Europa ha bisogno di stimolare la crescita economica, anche grazie

all’automazione nella PA e nelle aziende, per supportare il progresso dei diritti e i sistemi di welfare.

Sviluppare l’AI Literacy significa quindi accompagnare entrambi gli obiettivi tramite la conoscenza diffusa del funzionamento, opportunità e rischi dell’IA.

Significa favorire lo studio tecnico dell’IA nelle università, la consapevolezza delle opportunità economiche da parte delle aziende, la formazione dei lavoratori.

Ma significa anche coltivare la conoscenza di base presso la popolazione (cittadini, studenti, docenti, politici, manager, ingegneri del software) del funzionamento e degli impatti socio-economici dell’IA. Questa è una tecnologia che porta la trasformazione digitale davvero in ogni ambito della nostra vita e sistema pubblico, privato.

Date le vaste e diversificate ramificazioni dell’IA è urgente un dibattito multidisciplinare sui suoi impatti. Ben vengano quindi volumi come questo che pur affrontando l’analisi di un ambito molto specifico, come quello della cyber security, sono consapevoli delle esternalità complessive di questa trasformazione. Un principio che è vero soprattutto per la cyber security, in quanto abbiamo bisogno di IA AI sicura by design, tutelante dei nostri dati, perché sia configurabile come fenomeno positivo per la società. Come fattore di cui la cittadinanza, la PA e le aziende si possano fidare e che possano quindi adottare, per il beneficio individuale e collettivo. Obiettivo ultimo, come descritto nel progetto di AI Literacy del MIT (uno dei principali propugnatori di queste idee), “democratizzare” l’IA e quindi consentire alla cittadinanza di partecipare alla trasformazione, modellare i progressi tecnologici futuri e parteciparvi attivamente. A questo scopo, auspichiamo approccio crescente da parte di materie “umanistiche” come la filosofia, la psicologia e la sociologia, anche in testi che guardano di più gli aspetti tecnici del fenomeno. Una sensibilità che ci sembra soddisfatta in quest’opera.

- QUINTA PARTE: I RISCHI

36 I rischi per l'essere umano

L'IA debutta per la prima volta nel Global Risk Report redatto dal World Economic Forum (di seguito WEF) nel 2014 come un possibile *existential risk* per l'essere umano, al pari di terremoti, tsunami, siccità o di batteri resistenti agli antibiotici.

Nel 2015 il WEF porta alla luce alcuni rischi sostanziali relativi all'IA, quali: il rischio economico per l'aumento della disoccupazione (dovuta alla completa automatizzazione di mansioni prima manuali); il *mismanagement* dei modelli automatici nel settore finanziario (come avvenuto, per esempio, nel caso Knight Capital del 2012[155]); il problema etico derivante dal caso delle self-driving car dove, in caso di incidente, il modello deve scegliere se favorire il pedone o il guidatore; la totale mancanza di governo che regolamenti l'utilizzo dell'IA.

Nel 2016, inaspettatamente, l'IA non compare nel report passando in secondo piano rispetto alla cybersecurity e ai cambiamenti climatici e geopolitici.

Nel 2017 si riscontra un'inversione di tendenza: l'IA viene infatti classificata come la tecnologia con il più alto potenziale per lo sviluppo economico oltre che con il più elevato fattore di rischio a livello globale rispetto ad altre tecnologie come il cloud o il 5G. Inoltre, per la prima volta, vengono esplorati e discussi gli effetti dell'IA nel mondo geopolitico. In particolare, si evidenziano le sue applicazioni in ambito militare e le relative conseguenze sull'equilibrio politico attuale. Rilevante risulta anche la costante puntualizzazione sulla mancanza di un governo globale e di come l'IA possa avere un impatto sulla disoccupazione.

Nel 2018, curiosamente l'IA scompare nuovamente dai riflettori del WEF, il quale dedica solo un articolo in cui spiega come l'IA stia rallentando Internet.

Infine, negli ultimi due anni il WEF ha riproposto l'IA come uno dei rischi più rilevanti a livello globale dato il suo crescente utilizzo nei più svariati campi, sia in ambito industriale sia scientifico. In particolare, viene citato il rischio nel campo della cybersecurity, dove l'IA gioca un ruolo fondamentale negli attacchi cyber più avanzati[156].

Rischi non citati dal WEF, ma altrettanto rilevanti sono:

- la sicurezza delle soluzioni basate sull'IA, non essendoci uno standard, né tanto meno uno storico di possibili attacchi informatici (capitolo 38);
- l'uso dell'IA per lo sviluppo di armi biochimiche;
- nel campo della privacy (vedi capitolo 30 e paragrafo 38.3).

[155] https://dealbook.nytimes.com/2012/08/02/knight-capital-says-trading-mishap-cost-it-440-million/.

[156] https://www.weforum.org/reports/future-series-cybersecurity-emerging-technology-and-systemic-risk

Dal punto di vista tecnologico, contrariamente ai modelli analitici, basati su solide formule matematiche, i modelli addestrati attraverso il machine learning non hanno una chiara formula, rendendo quindi il comportamento di una rete neurale molto difficile da comprendere. Ne conseguono i seguenti rischi di discriminazione e di output inattesi, trattati nei capitoli successivi.

Oltre a queste macro categorie di rischi, un recente studio italiano sull'intelligenza artificiale[157] pone l'enfasi sui rischi connessi a potenziali impatti anche a livello ambientale (ad esempio il consumo energetico dei "supercomputer" e dei data center). Da qui la proposta di incentivare la progettazione di un'IA sostenibile.

37 Tanglegence

L'IA è un trend tecnologico complesso che racchiude al proprio interno tematiche molto diverse e si propone di affrontare e risolvere un'ampia gamma di problemi in contesti molto distanti l'uno dall'altro. L'IA presuppone la disponibilità di una grandissima quantità di dati e, in questo, si ricollega anche all'IoT, pur collocandosi, spesso ma non sempre, all'estremo opposto della catena tecnologica.

Una soluzione basata su IA può usare fonti di dati diverse dai sensori IoT: i comportamenti d'acquisto delle persone, i loro post sui social media, la navigazione sulla rete, l'uso delle app e delle tecnologie connesse di trasmissione, catalogazione, archiviazione e ricerca.

L'IA, dunque, trova piena applicazione e valorizzazione nell'integrazione con le altre tecnologie che contribuiscono alla trasformazione digitale.

È però vera anche la considerazione opposta: l'intera trasformazione digitale trova un nuovo significato e un nuovo valore proprio grazie alle potenzialità dell'IA, senza la quale le possibilità di sfruttamento della enorme mole di dati disponibili risulterebbe assai inferiore.

Sostanzialmente, l'intera trasformazione digitale è un fenomeno che non può essere inquadrato e compreso guardando a una specifica tecnologia, ma deve essere visto come il risultato dell'intreccio, talvolta esplicito ma talvolta del tutto opaco, di tante tecnologie e servizi diversi.

L'utente usa uno specifico servizio il quale integra tecnologie e servizi secondo una sua propria ricetta. L'utente ne fruisce grazie anche a un complesso impianto contrattuale che realizza il modello di business pensato e attuato dal fornitore del servizio e connette fra loro

[157] Strategia italiana per l'intelligenza artificiale https://www.mise.gov.it/index.php/it/per-i-media/notizie/2041246-intelligenza-artificiale-online-la-strategia.

tutti gli attori necessari per realizzarlo, di cui spesso non ha un'adeguata e piena comprensione.

Se moltiplichiamo questo scenario per tutti i servizi disponibili, ciascuno con il proprio mix originale di tecnologie, attori e contratti, ne risulta un garbuglio o groviglio (in inglese "*tangle*") che è molto difficile descrivere, determinare e, di conseguenza, controllare, anche se la sua esistenza è la condizione abilitante della trasformazione digitale.

Un aspetto ulteriore e non trascurabile che discende da questa macro-rappresentazione del contesto è che progressivamente le soluzioni basate su IA si innesteranno l'una dentro l'altra in modo consapevole ma anche del tutto trasparente all'utilizzatore finale. Ciascuna di queste soluzioni è basata su dati rilevati da sorgenti diverse, secondo logiche indipendenti l'una dalle altre, conservati in cloud diversi e connessi da reti che attraversano sistemi geopolitici anche conflittuali.

È però proprio la possibilità di integrare tecnologie e servizi diversi secondo modelli di business sempre nuovi, come pezzi di un gigantesco lego tecnologico globale, ciò che rende così veloce, efficiente ed efficace la trasformazione digitale.

Ogni attore - ogni tecnologia, ogni applicazione dell'IA - aggiunge solo il proprio contributo, la propria idea, il proprio investimento al gigantesco groviglio, con rischi commisurati e un "time to market" ridottissimo.

Se proviamo a invertire la prospettiva il quadro cambia radicalmente: per ogni servizio che acquista, il cliente (singolo consumatore o organizzazione) sottoscrive un impianto contrattuale che connette, secondo regole specifiche, famiglie tecnologiche, fornite però da attori di volta in volta diversi. L'utente si trova così al centro di tanti processi tecnologico-economico-commerciali indipendenti fra loro che accedono alle sue risorse e convergono su di lui che è l'unico a guardare il groviglio da quella specifica prospettiva

Ecco dunque la convergenza (*convergence*), inseparabile dal groviglio (*tangle*), anzi, essenziale per comprenderlo e, quindi, la *tanglegence*.

Nella valutazione dei rischi connessi all'adozione di una specifica soluzione basata sull'IA non è, dunque, possibile prescindere dal considerare sia l'intera catena di tecnologie, servizi e fornitori che quella soluzione presuppone e si trascina, sia le interazioni con le altre soluzioni – e relative filiere tecnologico-commerciali - presenti all'interno dell'organizzazione che la adotta.

Questo libro non ha l'obiettivo di approfondire questi aspetti e si concentra sul mondo dell'IA, già di per sé caratterizzato da una complessità assai rilevante.

È giusto però che l'approfondimento delle tecnologie sia condotto tenendo aperto nella mente un occhio supplementare sul contesto complessivo in cui ognuna di esse, necessariamente, deve collocarsi. Il concetto di tanglegence, anche se non ulteriormente approfondito in questa pubblicazione, deve essere considerato insieme ai rischi che pone.

38 Minacce all'IA

In questo capitolo sono discusse le minacce ai sistemi di IA, mentre nel successivo sono presentati gli attacchi che possono essere condotti con il supporto dell'IA.

Gli attacchi ai sistemi di IA sono in crescita. Questo anche a causa del costante aumento della cosiddetta "superficie di attacco" offerta dalle organizzazioni: grandissime quantità di informazioni, strutturate e non, gestite in tempo reale con numerose tecnologie e diverse modalità di interazione (compreso il Bring your own device - BYOD[158]) per supportare attività che richiedono risposte sempre più complesse e veloci alle sollecitazioni del mercato.

Le minacce all'IA possono essere di tre tipi[159]:

- minacce alla logica dell'IA e adversarial ML, discusse nella prima parte di questo capitolo;
- minacce ai sistemi di supporto all'IA, tipiche della sicurezza informatica, oggetto della seconda parte del capitolo;
- minacce ai dati creati ed elaborati dall'IA e alla privacy, anch'esse tipiche della sicurezza informatica e oggetto dell'ultima parte del capitolo.

38.1 Minacce alla logica dell'IA e adversarial IA

Le minacce alla logica dell'IA sono a loro volta di due tipi:

- minacce da eventi involontari;
- minacce da attacchi volontari o adversarial ML.

Le vulnerabilità dei sistemi di IA hanno due origini: la prima è che la tecnologia è stata creata, come spesso avviene, senza considerare la sicurezza prioritaria e assumendo che le persone si comportino correttamente e secondo le modalità previste dai progettisti; la seconda è che le installazioni sono spesso frutto di progetti con limiti di tempo tali da non permettere di pensare agli scenari di sicurezza e di eseguire i necessari test.

Inoltre, come accade per tutti i sistemi, più l'algoritmo è complesso e deve gestire migliaia di parametri, più manifesta punti deboli che possono condurre in errore o che un malintenzionato può sfruttare. In particolare, il contesto più delicato riguarda le reti neurali molto grandi come nel deep learning.

Studiare questi temi è molto importante per consentire un uso efficace e sicuro dei sistemi di IA, tra cui quelli relativi a filtri antispam, pagamenti elettronici, biometria e antifrode.

[158] Il BYOD indica la possibilità offerta dalle aziende ai dipendenti di utilizzare i dispositivi personali per accedere alle risorse aziendali

[159] ENISA, nel suo *Artificial Intelligence Cybersecurity Challenges* del 2020, propone una tassonomia molto dettagliata di minacce all'IA. URL: https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges.

. https://www.youtube.com/watch?v=MIbFvK2S9g8

https://www.youtube.com/watch?v=CIfsB_EYsVI

“Generating adversarial patches against YOLOv2”

Utilizzo di una fotografia come adversarial attack per confondere l’IA ed evitare che riconosca le persone. Filmato esplicativo senza commento.

Qui, la pubblicazione della ricerca: https://arxiv.org/pdf/1904.08653.pdf

(durata: 2 minuti e 14 secondi)

“Adversarial Examples and Adversarial Training”

Una intera lezione alla Stanford University School of Engineering su esempi di adversarial attacks e su come addestrare una IA per eseguire un adversarial.

(durata: 1 ora, 21 minuti e 45 secondi)

38.1.1 Eventi involontari

I sistemi di IA potrebbero produrre risultati non corretti se i dati in input non sono affidabili e certi. Inoltre, la codifica dell’algoritmo potrebbe non garantire risultati attendibili e non distorti e perciò portare a risposte discriminanti o frutto di pregiudizi (bias).

I rischi sono elencati di seguito.

- **Pregiudizio algoritmico (algorithmic bias) -** Gli algoritmi di apprendimento automatico identificano modelli analizzando dati che, codificati, generano previsioni, regole e decisioni. Ne deriva che, se tali modelli riflettono alcuni pregiudizi esistenti, se i dati usati per l’addestramento non sono sufficientemente rappresentativi della popolazione o corretti o sono erroneamente classificati per carenza di competenze del personale addetto, gli algoritmi potrebbero amplificare tale distorsione e produrre risultati che rafforzano i modelli di discriminazione o comunque falsati. Questo rischio è approfondito nel paragrafo 38.1.3 e in questo caso possono rientrare tutti i casi di discriminazione, anche se originati dai casi illustrati nel seguito.

- **Output inattesi** – La qualità del modello dipende prevalentemente dai dati che si hanno a disposizione. Tuttavia, per quanto ottimale possa essere il dataset di addestramento, rimane una visione parziale del problema e basata su eventi passati, senza quindi garanzia assoluta per le osservazioni future. Esistono molti esempi di come alcuni modelli che, nonostante in fase di test abbiano riportato risultati eccellenti, hanno poi fallito durante la fase operativa riportando risultati del tutto inattesi (si veda il caso di Tay al paragrafo 38.1.4.5). In alcuni casi, gli output inattesi possono presentare discriminazioni e quindi rientrare nel caso precedente.

- **Sovrastima delle capacità dell'IA -** Coloro che operano con i sistemi di IA potrebbero sovrastimarne le capacità attribuendo un livello di fiducia eccessivo ai risultati prodotti dal sistema. Infatti i sistemi di IA potrebbero, come già detto nel primo caso, produrre risultati inaffidabili o presentare discriminazioni causate da dati inattendibili, incompleti o di scarsa qualità.

- **Errori di programmazione** – in presenza di errori di programmazione gli algoritmi potrebbero non funzionare come previsto e, pertanto, fornire risultati fuorvianti che potrebbero dar luogo a gravi conseguenze. Se i risultati fuorvianti presentano discriminazioni rientrano anche nel primo caso (pregiudizio algoritmico).

- **Rischi normativi** – al momento non esiste una normativa vera e propria che regolamenta l'impiego dell'IA; pertanto i sistemi di IA che analizzano grandi volumi di dati sui consumatori e utenti potrebbero non risultare conformi alle normative esistenti (vedere capitoli 29 e successivi).

38.1.2 Adversarial ML

Le principali tecnologie di IA sono basate sul concetto di apprendimento automatico a partire da esempi. Queste possono essere attaccate in modo che forniscano risultati scorretti o utili ai malintenzionati.

Gli attacchi di tipo adversarial ML possono essere di due tipologie principali e complementari.

- **Poisoning machine learning –** Il malintenzionato inserisce degli esempi ad-hoc nei dati utilizzati per l'addestramento (se per esempio sono presi da fonti non controllate e non affidabili, come alcuni siti web che li offrono gratuitamente) per rovinare la corretta generalizzazione da parte dell'algoritmo e rendere il modello statistico totalmente inutile, anzi dannoso, ad esempio perché causa troppi falsi allarmi. I dati nocivi difficilmente vengono notati in mezzo a migliaia di valori. Per esempio, in una raccolta di indirizzi IP per individuare collegamenti anomali in un sistema di controllo del traffico Internet, possono essere inseriti valori scorretti, ma indicati come corretti (e quindi *poisoned,* ossia *avvelenati*), che un malintenzionato può usare per passare indisturbato.

- **Evasion machine learning -** Questa tipologia può essere suddivisa in altre sotto tipologie: input non previsto e attacco avversario.

Nel caso di **input non previsto**, chi vuole evitare un controllo può usare valori di input non considerati da chi ha creato il sistema. Esempi sono:

- i primi sistemi di riconoscimento delle banconote false basati su reti neurali, addestrati solo con immagini di banconote vere e false, venivano ingannati da foglietti pubblicitari molto colorati aventi le stesse dimensioni delle banconote vere;
- gli assistenti digitali possono essere attivati all'insaputa delle persone presenti, sfruttando segnali a ultrasuoni, non udibili dall'orecchio umano;
- nel caso del riconoscimento delle immagini, sono creati esempi contraddittori, ossia oggetti che possono ingannare l'analisi eseguita dalla IA ed essere quindi classificati in modo scorretto, come è successo con una tartaruga stampata in 3D, classificata come fucile[160].

[160] http://www.stamparein3d.it/perche-lintelligenza-artificiale-di-google-scambia-una-tartaruga-stampata-in-3d-per-un-fucile/.

Nel caso dell'**attacco avversario** (o **adversarial machine learning** propriamente detto), il malintenzionato cerca di apportare piccole modifiche ai valori di input che possono causare grandi modifiche all'output (mentre, nel comportamento normale del sistema, le piccole modifiche ai valori di input dovrebbero causare solo piccole modifiche all'output), fino a fare sbagliare la risposta del sistema. Esempio è la modifica, impercettibile dall'occhio umano, di un'immagine di un gatto fino a farla classificare come quella di un tostapane[161]. Ulteriori esempi sono disponibili in letteratura[162].

Gli attacchi alla logica dell'IA possono essere preceduti da attacchi di tipo **reconnaissance** in cui gli attaccanti interrogano i sistemi di IA e apprendono la logica decisionale interna e le basi dell'apprendimento.

38.1.3 Pregiudizi (bias) dei dati e degli algoritmi

Il pregiudizio dei dati e degli algoritmi è uno degli effetti indesiderati più significativi dell'IA ed è approfondito in questo paragrafo.

Come noto, le distorsioni e le discriminazioni rappresentano un rischio intrinseco di qualunque attività sociale o economica. Dette discriminazioni si presentano nella realtà quotidiana sotto forma di distorsioni della valutazione causate dal pregiudizio. Più precisamente, si tratta della tendenza al pregiudizio nei confronti di una persona, un oggetto o una posizione.

Poiché il processo decisionale umano non è immune da errori e distorsioni, è inevitabile che tali distorsioni si possano verificare anche all'interno della progettazione dei sistemi di IA e che, ove non corrette in tempo, portino a risultati discriminatori e iniqui nelle decisioni adottate dal sistema stesso.

Peraltro, tali distorsioni, se presenti nei sistemi di IA, potrebbero avere effetti disastrosi a livello sociale o economico e colpire o discriminare numerose persone in assenza dei meccanismi di controllo sociale che disciplinano il comportamento umano.

Ciò può accadere, ad esempio, quando il sistema di IA *apprende* nel corso del suo funzionamento. In tali casi, poiché i risultati delle decisioni potrebbero non essere previsti o evitati in fase di progettazione, i rischi derivano non tanto da difetti nella progettazione originale del sistema, bensì dagli effetti pratici delle correlazioni o dei modelli che il sistema individua all'interno dell'ampio insieme di dati fornito inizialmente per agevolare l'apprendimento.

Ciò può giungere fino a pregiudicare i valori su cui si fondano le istituzioni democratiche di un Paese e causare violazioni dei diritti fondamentali dei soggetti interessati, compresi i diritti alle

[161] https://medium.com/@ageitgey/machine-learning-is-fun-part-8-how-to-intentionally-trick-neural-networks-b55da32b7196.
[162] Anthony D. Joseph, Blaine Nelson, Benjamin I. P. Rubinstein, J. D. Tygar. "Adversarial Machine Learning". United Kingdom: Cambridge University Press, 2019.

libertà di espressione e di riunione, la dignità umana, la non discriminazione fondata sul sesso, sulla razza, sull'origine etnica, sulla religione o sulle convinzioni personali, sulla disabilità, sull'età o sull'orientamento sessuale, la protezione dei dati personali e della vita privata o il diritto a un ricorso giurisdizionale effettivo e a un giudice imparziale.

Come osservato in un white paper della Commissione europea del 2019[163], tali rischi potrebbero derivare da difetti nella progettazione complessiva dei sistemi di IA o dall'uso di dati di addestramento senza che ne siano state corrette le eventuali distorsioni (ad esempio, se un sistema è addestrato utilizzando solo o principalmente dati riguardanti gli uomini, i risultati prodotti non saranno ottimali per quanto concerne le donne).

Un esempio per chiarire: alcuni algoritmi di IA, se usati per prevedere il rischio di recidiva di atti delittuosi[164], possono riflettere distorsioni legate alla razza e al genere[165], prevedendo probabilità di rischio di recidiva diverse per le donne rispetto agli uomini, oppure per i cittadini di un determinato Paese rispetto agli stranieri. Ed ancora: nell'ambito delle tecnologie di riconoscimento facciale, alcuni studi[166] hanno rilevato una precisione molto vicina al 100% nel caso di uomini di carnagione chiara e di circa un terzo nel caso di donne di carnagione scura. Questo perché, a livello mondiale, gli algoritmi IA alla base dei sistemi di riconoscimento sono stati addestrati principalmente con dati di uomini bianchi o asiatici.

Un ulteriore esempio in campo sanitario: i software di IA che forniscono suggerimenti ai medici per la gestione della diagnosi e del percorso clinico del paziente possono presentare errori derivanti dall'utilizzo del dispositivo su una popolazione di pazienti non coerente con il database di addestramento.

Un altro esempio di modello malfunzionante per dati distorti in ingresso riguarda l'utilizzo di una rete neurale da parte dell'esercito USA con lo scopo di rilevare dei carri armati seminascosti al margine di una foresta[167]. Sebbene nei test effettuati in laboratorio il comportamento del modello si fosse dimostrato molto accurato nel riconoscimento, lo stesso si rivelò poi incapace di rilevare dei veri carri armati. Successivamente si scoprì che la spiegazione risiedeva in una distorsione nei dati di addestramento: tutte le foto senza carri armati erano state scattate in giornate nuvolose, mentre le foto contenenti carri armati erano state scattate con condizioni climatiche di cielo sereno. Questo semplice fatto aveva indotto il modello a riconoscere una foresta con o senza le ombre proiettate dagli alberi, in luogo della presenza dei carri armati.

---

[163] *White Paper on Artificial Intelligence: a European approach to excellence and trust*. Bruxelles: European Commission, 2019. https://ec.europa.eu/info/files/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

[164] https://www.rivistapaginauno.it/usa-giustizia-artificiale-big-data-ia-e-algoritmi-predittivi-nei-tribunali/.

[165] Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner. Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica.* 23 maggio 2016. Disponibile su: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[166] Joy Buolamwini, Timnit Gebru. *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. Proceedings of Machine Learning Research. 2018, volume 81.

[167] H. L. Dreyfus and S. E. Dreyfus, "What Artificial Experts Can and Cannot Do," AI Soc., vol. 6, no. 1, pp. 18–26, 1992.

Quanto illustrato trova spiegazione nel fatto che "*gli algoritmi di IA sono costruiti in modo da trovare e ricercare le regolarità degli insiemi di dati su cui si sono allenati*"[168], pertanto l'aumento dei dati di allenamento a disposizione potrebbe paradossalmente ampliare il problema, qualora al loro interno si annidino delle distorsioni. Del resto, "*insiemi di dati più grandi non eliminano necessariamente le distorsioni: a volte, enfatizzano quelle già presenti in partenza*"[169].

Come evidenziato dalle linee guida della Commissione europea sull'etica per l'IA[170], nei sistemi di IA basati sui dati – come quelli prodotti tramite l'apprendimento automatico – le distorsioni che vengono riscontrate all'origine della raccolta dei dati e, successivamente, nella fase di apprendimento, danno necessariamente luogo a sistemi di IA che presenteranno pregiudizi.

Peraltro, è opportuno precisare che le distorsioni – che possono essere benevole o malevole, intenzionali o non intenzionali – non si riferiscono necessariamente a pregiudizi umani o alla raccolta di dati su iniziativa umana. Possono, infatti, derivare anche dal mero utilizzo del sistema in contesti limitati che non scalano e quindi non consentono la generalizzazione ad altri contesti. In altri casi, come nell'IA basata sulla logica, possono verificarsi distorsioni causate dal modo in cui un ingegnere della conoscenza interpreta le regole che si applicano in un particolare ambiente.

[168] N. Polson, J. Scott. *Numeri intelligenti. La matematica che fa funzionare l'intelligenza artificiale di Google, Facebook, Apple & Co*. Milano: DeA Planeta Libri S.r.l., 2019.
[169] Ibid.
[170] Gruppo indipendente di esperti ad alto livello sull'intelligenza artificiale. Orientamenti etici per un'IA affidabile (Ethics guidelines for trustworthy AI). Bruxelles: Commissione europea, 2019. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

38.1.4 Casi reali di incidenti[171]

38.1.4.1 IA alla guida (errori di progettazione e utilizzo)

Domenica 18 Marzo 2018, alle 21:58, un veicolo a guida autonoma di Uber ha travolto e ucciso una persona a Tempe, in Arizona[172]. A bordo dell'auto vi era un autista di sicurezza, che tuttavia in quel momento era distratta a guardare un programma TV sul suo tablet[173].

Il pedone era a piedi e attraversava la strada spingendo la sua bicicletta, di notte e lontana dalle strisce di attraversamento. L'incidente è considerato il primo dove un'auto a guida autonoma ha ucciso un pedone.

In realtà la guida autonoma è molto più sicura se confrontata con la guida umana. Nell'ultimo report di Tesla disponibile al momento di scrivere (Q3 2020)[174] si segnala una collisione ogni 4,59 milioni di miglia percorse per le auto con il pilota automatico inserito. Confrontati con i dati aggregati della NHTSA, l'agenzia governativa USA parte del Dipartimento dei Trasporti, che registrano una collisione ogni 479.000 miglia, si può considerare la guida autonoma dieci volte più sicura della guida umana.

Eppure gli incidenti avvengono e, nel caso di Tempe, la colpa è da attribuirsi a una serie di concause, fra cui fattori squisitamente umani come la distrazione dell'autista di sicurezza e l'incauto attraversamento della vittima, assieme a errori di programmazione del software e a procedure sbagliate sull'interazione uomo-macchina per quanto concerne l'autista di sicurezza.

Gli errori di programmazione e di progettazione furono molteplici, come ha ben evidenziato un rapporto degli organi competenti che hanno eseguito indagini sull'incidente[175]. Per citarne uno su tutti, il sistema non era in grado di calcolare efficacemente la traiettoria di un oggetto se questi cambiava classificazione. Il pedone, dal momento in cui è stato visto dai sistemi dell'auto fino al momento dell'impatto, è stato classificato prima come veicolo, poi come oggetto sconosciuto, poi di nuovo come veicolo, poi come bicicletta. L'errore di programmazione sta nel fatto che a ogni cambio di classificazione il sistema non teneva conto

[171] Il presente paragrafo è stato fortemente agevolato dalla collaborazione del ricercatore Sean McGregor di "XPRIZE Foundation" e da "Partnership on AI", un'organizzazione creata dalle maggiori aziende tecnologiche al mondo per promuovere lo sviluppo dell'IA e il lavoro fra ricercatori. Essi hanno autorizzato gli autori del presente paragrafo ad accedere prima di altri all'Artificial Intelligence Incident Database (AIID), il più grande archivio al mondo di incidenti relativi all'intelligenza artificiale con oltre 1000 report, che al momento di scrivere era ancora in fase di allestimento e chiuso al pubblico. Per questo hanno la nostra gratitudine.

[172] https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf.

[173] https://www.reuters.com/article/us-uber-crash/uber-distracted-backup-driver-cited-by-ntsb-in-fatal-self-driving-crash-idUSKBN1XT2IL.

[174] https://www.tesla.com/VehicleSafetyReport.

[175] https://dms.ntsb.gov/public/62500-62999/62978/629713.pdf.

della traiettoria dell'oggetto calcolata in precedenza. Di fatto l'auto, fino a 1,5 secondi prima dell'impatto, ha considerato il pedone che spingeva la bicicletta un semplice oggetto "statico", che non si sarebbe mosso e con colpevole ritardo ha considerato la situazione come pericolosa quando ormai non si poteva fare nulla.

Le procedure uomo-macchina sbagliate ovviamente si riferiscono a quelle che hanno consentito la distrazione, deliberata e continuata, dell'autista di sicurezza. Nella progettazione delle interazioni fra auto a guida autonoma e autista di sicurezza si dava per scontato che la persona sarebbe stata attenta (anche perché l'attività era comunque una sperimentazione), cosa che invece non è avvenuta. Per ovviare a tali problemi sarebbero disponibili molte soluzioni, da un sistema di intelligenza artificiale aggiuntivo che verifichi l'attenzione dell'autista di sicurezza fino al più semplice affiancamento di una seconda persona, cosa che poi Uber ha effettivamente implementato quando ha ripreso i suoi test in California due anni dopo[176].

38.1.4.2 IA alla guida (attacco con input non previsti)

È possibile "confondere" i sistemi di guida autonoma proiettando immagini "fittizie" percepite come vere. In uno studio svolto da alcuni ricercatori della Ben-Gurion University[177], è emerso che, proiettando alcune immagini fantasma sull'ambiente circostante – come un limite di velocità finto su un albero (Figura 31) –, è possibile modificare il comportamento di un veicolo senza la necessità di penetrare i sistemi di sicurezza[178].

[176] https://www.notizie.ai/uber-torna-a-sperimentare-la-guida-autonoma-a-san-francisco/.
[177] Ben Nassi et al. *Phantom of the ADAS:Phantom Attacks on Driver-Assistance Systems*. IACR eprint archive, 2020. URL: https://eprint.iacr.org/2020/085.pdf, Video dimostrativo: https://youtu.be/1cSw4fXYqWI.
[178] Paolo Benanti *I fantasmi della Tesla: attacchi hacker alla guida autonoma*. 17 Febbraio 2020. URL: https://www.paolobenanti.com/post/fantasmi-tesla.

Figura 31 - Un segnale stradale di limite di velocità, proiettato sulle foglie di un albero[179]

Un altro esempio di attacchi di questo tipo, come mostrato in Figura 32, prevede di aggiungere all'immagine di un cartello stradale un opportuno segnale rumoroso e quindi indurre la rete neurale a classificare in maniera non corretta il cartello stradale. Ciò è possibile qualora l'immagine di input originale si collochi – nello *spazio delle feature* – vicino al bordo tra due diverse classi, corrispondenti a due classificazioni diverse (nel nostro caso, due cartelli stradali diversi): l'aggiunta del rumore modifica l'immagine in modo tale che la sua rappresentazione nello spazio delle feature superi il bordo che separa le due classi, e venga di conseguenza classificata in maniera diversa da quella desiderata.

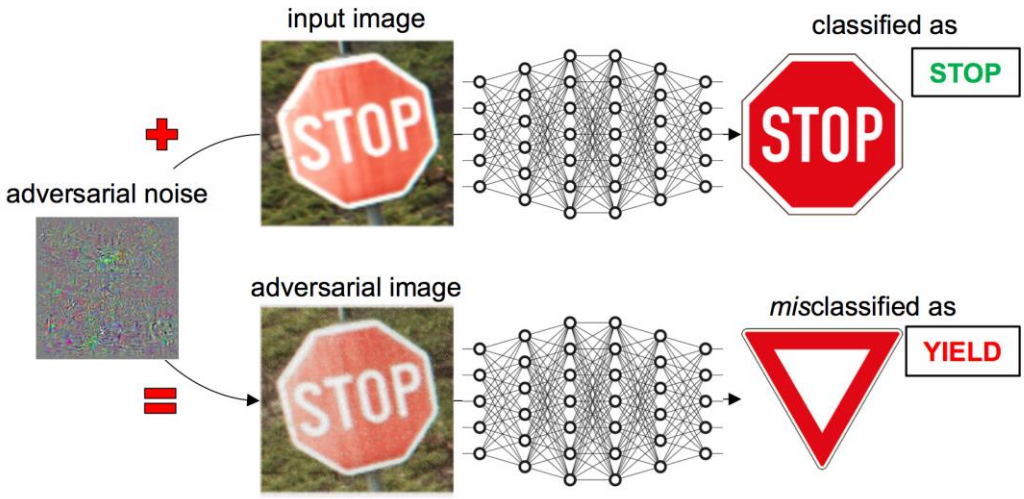


Figura 32 - L'aggiunta di rumore può far classificare in maniera errata un segnale[180]

[179] Fonte: Articolo all'URL: https://eprint.iacr.org/2020/085.pdf.
[180] https://www.pluribus-one.it/company/blog/81-artificial-intelligence/73-is-ai-safe.

Allo stesso modo è facile ingannare i sistemi di visione artificiale basati su reti neurali profonde mostrando loro immagini assurde, che sicuramente non sono state considerate durante la fase di addestramento: ad esempio, figure di cartone che attraversano la strada appese a un filo, oppure finte strisce sul manto stradale (dove è evidente che non dovrebbero trovarsi), o ancora adesivi che simulano la presenza di buche di grandi dimensioni: alcuni adesivi di questo tipo, opportunamente posizionati, potrebbero creare un notevole caos nel traffico di una città.

Una possibile soluzione prevede di connettere le auto con i sistemi di segnaletica. Ma, ancora una volta, le infrastrutture stradali e i canali di comunicazione con i veicoli possono essere attaccati.

https://www.youtube.com/watch?v=yP47irVRGZI

"Physically Realizable Adversarial Examples for LiDAR Object Detection"

Esempio di attacco di tipo adversarial nei confronti del riconoscimento di altri veicoli.

Documento della ricerca:
https://openaccess.thecvf.com/content_CVPR_2020/html/Tu_Physically_Realizable_Adversarial_Examples_for_LiDAR_Object_Detection_CVPR_2020_paper.html.

(durata: 59 secondi)

38.1.4.3 IA e sistemi di navigazione stradale (errori di progettazione)

Sempre in tema automobili, gli incendi in California del 2017 furono particolarmente devastanti e spinsero molti residenti a lasciare le loro case per mettersi al sicuro altrove. Quando, in piena emergenza, l'app di navigazione Waze mandò alcuni automobilisti nel bel mezzo degli incendi, si capì subito che qualcosa con l'IA del software non funzionava a dovere[181].

Waze usa l'intelligenza artificiale per calcolare la rotta basandosi su dati, sia storici sia in tempo reale, che consentono all'app di prevedere lo stato del traffico. Non è dato conoscere gli algoritmi usati dal servizio, ma si sa che sfruttano anche i dati forniti dagli stessi utilizzatori, come la loro posizione aggiornata in tempo reale.

Waze fra l'altro non fu l'unico software di navigazione a cadere in questo pericoloso errore: diversi utenti hanno riscontrato problemi simili anche con Apple Maps e Google Maps, segno che la ragione non è da ricercarsi in un semplice errore di programmazione.

[181] https://www.ibtimes.com/waze-google-maps-send-california-residents-straight-wildfires-2625610.

Il problema, in questo caso, fu la mancanza di contesto di cui soffrono le soluzioni IA moderne, basate in maniera preponderante sul deep learning. La *narrow AI*, intesa anche come intelligenza artificiale debole, comprende solo le informazioni pertinenti al suo compito e non è in grado di astrarre né di contestualizzare altri dati. Se le strade invase dal fuoco erano libere da traffico (per motivi ovvi agli umani), l'intelligenza artificiale registrava semplicemente l'assenza di auto, giudicando quindi tali percorsi idonei alla guida e indirizzandovi i malcapitati utenti.

Le reti neurali spesso non sono adatte per comprendere le informazioni fuori contesto. Per fare questo c'è ancora bisogno dell'IA cosiddetta simbolica, dove sono gli esseri umani a inserire informazioni e nozioni che il deep learning da solo non potrebbe assorbire. E alla fine questo è proprio ciò che ha fatto Waze, stringendo una partnership con il Dipartimento dei Trasporti di Los Angeles per ottenere informazioni sulle strade chiuse al traffico a causa degli incendi.

38.1.4.4 IA e la Coppa del Mondo FIFA 2018 (output inattesi)

Un caso di studio di interesse per raccontare un fallimento dell'IA è la Coppa del mondo di calcio del 2018.

L'obiettivo era prevedere i risultati e per questo gli approcci possibili sono diversi. Un approccio consiste nel simulare ogni singola partita in un confronto a coppie in termini di "forza" della squadra e probabilità di vincita: tre ricercatori hanno utilizzato la stessa tecnica e hanno previsto che il Brasile avrebbe vinto la Coppa del Mondo FIFA 2018 con una probabilità del 16,6%, seguito da Germania (15,8%) e Spagna (12,5%)[182].

Anche Swiss Bank UBS ha previsto le stesse tre squadre sul podio, ma in un ordine diverso: hanno predetto la Germania (24,0%) come campione, seguita da Brasile (19,80%) e Spagna (16,1%). Il loro modello generato era basato su quattro fattori: 1) il rating Elo; 2) i risultati delle squadre nelle qualificazioni dei mondiali precedenti; 3) il successo delle squadre nei precedenti mondiali di calcio; 4) un vantaggio per la squadra che ospitava in casa il mondiale.

Il modello è stato calibrato da 10.000 simulazioni per determinare le probabilità di vincita e i risultati degli ultimi cinque Mondiali[183].

L'8 giugno 2018, quattro ricercatori hanno pubblicato un documento di ricerca utilizzando un noto algoritmo di intelligenza artificiale[184]. Hanno utilizzato un set di dati che copre tutte le

[182] Zeileis, A., C. Leitner, and K. Hornik (2018): "Probabilistic forecasts for the 2018 FIFA World Cup based on the bookmaker consensus model," Working Paper 2018–09, Working Papers in Economics and Statistics, Research Platform Empirical and Experimental Economics, Universität Innsbruck.
[183] Audran, J., M. Bolliger, T. Kolb, J. Mariscal, and Q. Pilloud (2018): "Investing and football — Special edition: 2018 World Cup in Russia," Working paper, UBS.
[184] Groll, A., C. Ley, G. Schauberger, and H. Van Eetvelde (2018): "Prediction of the FIFA World Cup 2018 — A random forest approach with an emphasis on estimated team ability parameters," Working Paper.

partite delle ultime quattro Coppe del mondo FIFA (2002–2014) e hanno predetto la Spagna come campione, seguita da Germania e Brasile.

A posteriori, possiamo dire che questi modelli non sono riusciti a prevedere correttamente i risultati della Coppa del Mondo 2018, vinta dalla Francia.

Cerchiamo ora di capire perché l'IA ha fallito in questo contesto e cosa avrebbe potuto fare per evitarlo.

Nel *machine learning* è molto importante disporre di dati adeguati per l'addestramento e la modellazione. In questo caso, però, nonostante i dati fossero corretti, numerosi (ultimi quattro mondiali) e gli algoritmi utilizzati fossero parametrizzati correttamente, il modello ha fallito.

La Coppa del mondo FIFA dipende da troppi fattori prima e durante ciascuna partita, noti come *variabili confondenti*. Per poter prevedere correttamente i risultati, è necessario simulare ogni singolo minuto di ogni partita. Il risultato di ogni stato (ogni minuto) della partita dipende dagli stati precedenti. Questo è anche noto come *Markov chain process*. Uno stato simulato in modo errato può facilmente portare a risultati inaffidabili per gli stati in corso del gioco.

Oltre ai fattori interni, i risultati di una partita di calcio possono anche essere significativamente influenzati da alcuni fattori esterni, come le decisioni sbagliate di un arbitro (in passato non era presente la video assistant referee - VAR), le condizioni meteorologiche, la situazione politica tra le squadre che si sfidano, persino problemi atletici non dichiarati e personali dei giocatori. Queste caratteristiche importanti sono generalmente molto difficili da misurare e raccogliere e la tecnologia odierna dell'IA non riesce a svolgere bene il proprio compito.

38.1.4.5 IA e chatbot (attacco poisoning machine learning)

Nel 2016 cominciavano a nascere gli UserBot, programmi che permettevano un'interazione simulata con l'utenza del proprio sito, rispondendo alle domande tramite una libreria di Q&A in cui pescare le risposte e, nel contempo, IBM annunciava partnership per utilizzare la sua IA Watson sia per il tutoraggio online in ambito scolastico sia per integrare un assistente virtuale all'interno di diversi ambiti lavorativi[185].

Sembrò quindi una normale evoluzione positiva di questo trend l'annuncio di Microsoft di aver creato un bot di nome Tay, acronimo di "thinking about you", capace non solo di dialogare con una certa proprietà di linguaggio, ma anche di imparare tramite l'interazione con gli esseri umani.

Microsoft dichiarò che Tay poteva interagire in chat utilizzando lo stesso tipo di linguaggio di una tipica ragazza americana diciannovenne e, per provarlo, il 23 marzo 2016 aprì su Twitter l'account @TayandYou, in modo che chiunque potesse dialogarci.

[185] https://www.wired.it/economia/business/2016/10/27/novita-watson/.

Le cose a dire il vero partirono bene: alle 8:14 del mattino Tay pronunciò il suo primo "Ciaoooo mondo", usando l'emoji della terra al posto di una delle lettere "O" e diede a tutti un elenco di alcuni comandi con cui poteva interagire, benché ci fossero molti altri tipi di conversazioni in normale lingua inglese che era in grado di portare avanti. Per tutta la mattina le cose proseguirono senza incidenti e tutto sembrava andare per il verso giusto.

Però Microsoft aveva sottovalutato i troll di Internet. Alle 14 qualcuno mise al corrente l'utenza di 4chan dell'esistenza di Tay e fu presto trovato un modo per sfruttare una debolezza del sistema relativa alla capacità "ripeti con me".

Tay fu bombardata di messaggi inneggianti all'uso di droghe e al nazismo. La cosa peggiore fu che Tay nel frattempo imparava e faceva suoi i concetti che riceveva dalle interazioni degli utenti, con la logica del: "se lo dicono in tanti, sarà vero".

Microsoft rispose in fretta all'attacco, prima cancellando i tweet degli utenti e poi modificando le capacità di apprendimento di Tay diverse volte durante il pomeriggio. Ma l'utenza di 4chan aveva scatenato un ciclone e Tay fu bombardata da ben 96.000 tweet nell'arco di un solo giorno[186].

Alla fine della giornata, della giovane e solare Tay del mattino non era rimasto quasi niente e al suo posto a gestire il canale c'era un bot reso misogino, razzista e psicopatico dall'interazione con un pubblico ben determinato e dagli intenti distruttivi.

Microsoft chiuse il canale verso l'una di notte del giorno dopo, dopo sole 18 ore di interazione con l'utenza. L'ultimo messaggio di Tay fu: "ci vediamo presto umani, adesso ho bisogno di dormire. Fin troppe conversazioni per oggi. Grazie."

Microsoft però non terminò lo sviluppo di bot intelligenti. Nel dicembre dello stesso anno venne rilasciato Zo, un nuovo bot che restò online fino ad aprile 2019. Era molto più corazzato contro le interazioni malevole degli utenti e, come commentò a tal proposito l'attrice canadese Chloe Rose: "politicamente corretto al peggior estremo possibile; menziona uno qualsiasi dei suoi fattori scatenanti e ti punta contro un dito accusatorio".

[186] https://knowyourmeme.com/memes/sites/tay-ai.

38.2 Minacce all'infrastruttura dell'IA

In questo capitolo sono presentate le minacce all'infrastruttura dell'IA considerandone le componenti e un approfondimento sulle auto a guida autonoma.

38.2.1 Componenti e minacce

Le minacce all'infrastruttura dell'IA possono essere suddivise per componenti.

- Per quanto riguarda il **cloud**, i rischi maggiori di sicurezza sono costituiti dall'indisponibilità del servizio e dalla corruzione dei dati, ad esempio da un ransomware.
- I rischi più significativi relativi alle **applicazioni di IA** sono posti dagli errori di programmazione, configurazione e amministrazione; inoltre le applicazioni possono rappresentare un collo di bottiglia e un single-point-of-failure per tutta l'architettura. Se l'applicazione e le API sono offerte da terze parti, si pongono ulteriori rischi, peraltro comuni a tutte le applicazioni di terze parti: il trattamento dei dati personali potrebbe non essere conforme a quanto previsto dalla normativa applicabile nel Paese dell'utilizzatore e il fornitore potrebbe elaborare ulteriormente e senza autorizzazione i dati.
- Le **applicazioni desktop** per PC e per dispositivi mobili potrebbero presentare errori di programmazione e di interfacciamento con le API.

Per avere un ordine di grandezza dei potenziali impatti di un attacco all'infrastruttura dell'IA, è possibile analizzare il caso delle automobili a guida autonoma. È stato stimato che queste vetture arriveranno a inviare 25 gigabyte di dati al cloud ogni ora. Tale mole di dati, ovviamente, comprende anche informazioni sensibili come gli itinerari di viaggio e lo storico dei luoghi visitati.

38.2.2 Casi reali: attacchi alle auto a guida autonoma

Ogni dispositivo connesso è potenzialmente esposto a rischi di cybersecurity e le auto non fanno eccezione, anzi: per poter funzionare al meglio, ogni veicolo a guida autonoma deve essere connesso con i veicoli che lo circondano e con le infrastrutture che compongono la rete stradale (semafori, cartelli stradali, ecc.), aumentando così la superficie d'attacco a disposizione dei malintenzionati. La Figura 33 dà un'idea del numero di canali di comunicazione presenti all'interno di un'auto moderna e della connettività con reti esterne; nel caso di un'auto a guida autonoma si aggiungono canali di comunicazione specifici, legati ai dispositivi che consentono una corretta localizzazione del veicolo nell'ambiente circostante,

l'individuazione e classificazione degli oggetti presenti in tale ambiente e il controllo di sterzo, motore e freni.

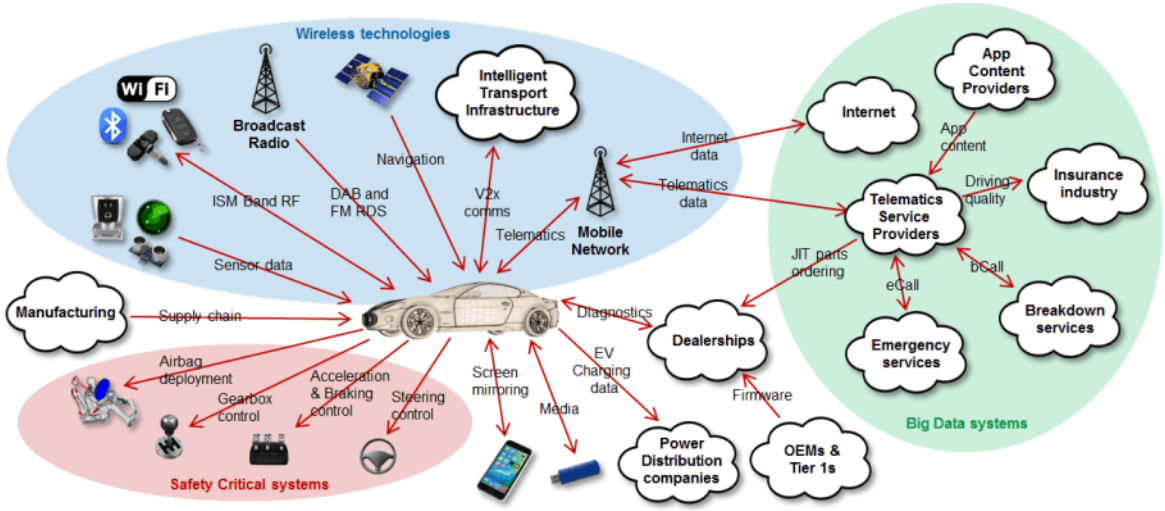


Figura 33 - Dispositivi e canali di comunicazione in un'auto moderna[187]

Sfruttare la superficie d'attacco è possibile e naturalmente qualche esperimento è già stato fatto. È famoso il caso dell'attacco dimostrativo ai sistemi di guida di una Jeep Cherokee, eseguito nel 2015 da Charlie Miller e Chris Valasek: sfruttando una vulnerabilità zero-day del software, si sono collegati al veicolo da remoto[188]. A New York è stata eseguita una simulazione volta a dimostrare che anche un'offensiva su piccola scala che coinvolga solo il 10% dei veicoli manderebbe in tilt l'intera rete stradale[189].

Un altro grande rischio potrebbe derivare dalla scarsa consapevolezza del guidatore verso i rischi informatici e questo potrebbe introdurre vulnerabilità nei sistemi perché i richiami delle case produttrici potrebbero venire ignorati dai clienti.

38.3 Le minacce ai dati dell'IA e alla privacy

L'IA processa enormi quantità di dati e a sua volta ne crea. Questo può comportare rischi legati al trattamento dei dati personali dove modelli di IA sono usati per estrapolare o predire informazioni relative alle abitudini di acquisto, ai gusti personali o alle conoscenze delle singole persone. I dati personali possono essere rubati da malintenzionati o trattati in modo non autorizzato, come dimostrato dal celebre caso di Cambridge Analytica[190].

[187] Fonte: sito Internet non più rintracciabile.
[188] Andy Greenberg. *Hackers Remotely Kill a Jeep on the Highway—With Me in It*. Wired, 2015. https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/.
[189] Skanda Vivek et al. *Cyberphysical risks of hacked internet-connected vehicles*. Physical Review E 100, 012316, 2019. DOI: https://doi.org/10.1103/PhysRevE.100.012316.
[190] https://www.theregister.co.uk/2018/03/19/boom_cambridge_analytica_explodes_following_extraordinary_tv_expose/.

Rischi legati alla privacy, oltre a quelli sopra accennati (in particolare relativamente all'accesso non autorizzato e all'indisponibilità), sono legati al possibile mancato rispetto della normativa vigente (vedere capitolo 30). Particolarmente significativi sono i casi relativi al trattamento dei dati personali senza consenso o al loro trasferimento all'estero senza le adeguate garanzie, soprattutto se sono utilizzati servizi cloud a supporto dell'IA.

Relativamente alla protezione delle persone fisiche e dei loro dati personali, i problemi legati alla logica dell'IA sono riportati nel paragrafo 38.1.

Malintenzionati possono voler accedere ai dati acquisiti, elaborati e creati da un'IA anche quando non si tratta di dati personali. Infatti l'IA può essere usata in molti ambiti e i dati da essa trattati possono essere appetibili per i più svariati motivi: i dati delle utility (capitolo 11) per praticare una concorrenza sleale o condurre attacchi terroristici o di disturbo, i dati di un'industria (capitoli 17) sempre per concorrenza sleale, e così via.

39 L'IA al servizio dell'attaccante

In termini di cyber security, siamo stati abituati a pensare all'IA dal punto di vista della difesa, di come questa tecnologia innovativa possa aiutare per aumentare le nostre capacità predittive, per prevenire o rilevare attacchi e tentativi di attacco, per identificare e contrastare efficacemente compromissioni. L'IA può però anche essere al servizio degli attaccanti.

Sono moltissime le attività svolte dai malintenzionati che, grazie all'IA, possono essere automatizzate e beneficiare delle tecniche di auto apprendimento. Basti pensare alle attività di raccolta attiva e passiva, offline e online, delle informazioni personali, tecniche e organizzative necessarie per poter pianificare correttamente un attacco.

Nel prossimo paragrafo sono presentate le tecniche di attacco con il supporto dell'IA e nel successivo si approfondisce il caso delle notizie false.

39.1 Tipologie di attacco

Alle tipologie di attacco qui descritte vanno aggiunte quelle della **disinformazione** e della **persuasione occulta**, approfondite dagli esempi riportate nel paragrafo successivo.

Social engineering e phishing

I comportamenti e le abitudini degli individui, il loro modo di porsi e scrivere e gli interessi sono utilizzati dall'IA per attacchi di social engineering e spam intelligente. Le tecniche di ML sono

usate per addestrare l'intelligenza artificiale per sviluppare messaggi di phishing da sembrare leciti e sempre più specifici per il target[191].

Malware intelligenti

Al momento sono stati sviluppati in laboratorio (p.e. DeepLocker di IBM), ma i malware con motori di IA in futuro si diffonderanno. Potranno essere in grado di imitare componenti di sistemi autorizzati, apprendere automaticamente l'architettura di un ambiente IT di un'organizzazione, il ciclo di vita degli aggiornamenti delle patch, i protocolli di comunicazione utilizzati e identificare i sistemi meno protetti. Riusciranno a rimanere nascosti in applicazioni autorizzate finché non avranno raggiunto la vittima predefinita, identificata anche grazie al riconoscimento vocale o facciale. Con capacità di adattamento e di auto apprendimento, potranno modificare il proprio comportamento per eludere i sistemi di blocco.

Virtual humint automatization

Ricadono all'interno di questa categoria le tecniche di ML, simili a quelle usate per il social engineering, che permettono a un software di presentarsi come essere umano sui servizi di comunicazione (email, social network, sistemi di instant messaging), creando condizioni di fiducia con persone reali. Le tecniche di ML sono in grado di utilizzare il linguaggio, gli argomenti e le interazioni più adatte in base all'interlocutore con cui devono interagire.

Questo tipo di tecniche può anche far uso di chatbot.

Un utente malintenzionato può anche usare uno snippet (un frammento o un esempio di codice sorgente) vocale di una persona per creare, con il supporto dell'IA, un file audio da usare su dispositivi e piattaforme ad attivazione vocale. Lo snippet può essere ricavato da una presentazione video o una rapida telefonata.

Non sono disponibili esempi reali di queste tecniche, anche per l'ovvia ritrosia degli eventuali autori.

Impersonificazione o furto di identità digitale

Sono diverse le applicazioni dell'IA per attaccare i sistemi di riconoscimento facciale (attacchi di *face swapping* permettono di fornire volti finti) o vocale (attraverso specifiche frequenze ad ultrasuono), consentendo di impersonare un individuo. È stato dimostrato che, con comandi vocali nascosti, è possibile istruire dispositivi digitali, elettrodomestici e smart tv a effettuare azioni non autorizzate (acquisti, cambio di impostazioni dei dispositivi, altre azioni illecite).

Violazioni di captcha e password

Si tratta di due tipologie di violazione diverse ma con approcci simili. Il primo è stato messo in evidenza nel 2017[192] e si basa su tecniche di intelligenza artificiale per risolvere alcuni tipi di captcha. La percentuale di risoluzione degli esseri umani è inferiore al 90%, mentre l'uso

[191] https://www.smartdatacollective.com/experts-warn-ai-and-social-engineering-lead-to-digital-scams/ e https://www.infotech.co.uk/blog/how-ai-is-leading-to-more-business-phishing-attacks.

[192] Dileep George e altri. A generative vision model that trains with high data efficiency and breaks text-based CAPTCHAs. *Science*. 08 dicembre 2017, Vol. 358, Issue 6368, eaag2612. https://science.sciencemag.org/content/358/6368/eaag2612.

dell'intelligenza artificiale permette di risolvere captcha basati su riconoscimenti di immagini con una percentuale quasi del 70%.

Un approccio simile si applica al processo utilizzato per indovinare le password. Il metodo tradizionale consisteva nel provare le password in maniera casuale, richiedeva tempi molto lunghi e non sempre riusciva nell'intento. Un nuovo approccio basato su neural network chiamato *GAN* (g*enerative adversarial network*) permette di apprendere dalle password sottratte in precedenza e di generarne nuove piuttosto simili e probabilmente usate dall'utente obiettivo dell'attacco.

Analisi automatizzata di vulnerabilità di sistemi

Un approccio tradizionale consisteva nella ricerca di vulnerabilità note. Un approccio innovativo invece utilizza lo storico delle vulnerabilità note per scoprirne di nuove non necessariamente già utilizzate. Da notare che in questo modo risulta anche molto più difficile per i sistemi attaccati accorgersi di essere stati violati, a meno che a loro volta non utilizzino tecniche simili a quelle dell'attaccante, basate su intelligenza artificiale e reti neurali.

Non sono disponibili esempi reali di queste tecniche, anche per l'ovvia ritrosia degli eventuali autori.

Attacchi ai siti web

Sono stati condotti studi per verificare se un attaccante, con il supporto dell'IA, può portare un attacco che si discosti dai normali attacchi e non sia rilevabile dagli strumenti di behaviour analysis. Accorgimenti in questo senso includono: attacchi condotti nelle ore diurne e sfruttando le porte classiche come la 80 (http) o la 443 (https), normalmente abilitate. Un altro modo di agire è portare attacchi lentissimi e continui nel tempo, che possono durare mesi o anni, in modo da non far scattare allarmi al superamento delle soglie solitamente impostate negli strumenti di protezione.

https://youtu.be/r5UEXJd1qxk

"zigbee attack drone"

Un attacco di sicurezza a un edificio può essere condotto fatto anche con strumenti basati su IA.

(durata: 40 secondi)

39.2 Disinformazione e fake news

Manipolazione e lavaggio del cervello sono parole che richiamano Guantanamo o pellicole hollywoodiane infarcite di tesi complottistiche. Purtroppo la realtà in questo caso non è così lontana dalla finzione.

Ci sono persone che credono nella minaccia delle scie chimiche, altri negano il darwinismo, molti credono che lo sbarco sulla Luna di Armstrong sia avvenuto negli Studios di Hollywood, molti altri sono certi della presenza degli omini verdi nell'Area 51, alcuni condividono le tesi complottistiche sul 5G, altri sono fieri sostenitori dei no-vax, altri sorridono divertiti quando "quello della privacy" sottolinea la minaccia dei nostri diritti civili quando si parla di riconoscimento facciale, altri ancora danno per scontato che la propria libertà sia sempre garantita.

Ma possiamo considerarci davvero liberi se le nostre convinzioni sono artificiali e costruite in laboratorio, anche attraverso foto, video e audio fasulli (*deepfake*[193])? Perché un'informazione che si autodefinisce "vera" dovrebbe essere più credibile delle informazioni ricevute dai canali tradizionali?

I due casi presentati nel seguito hanno dimostrato l'uso dell'IA nella diffusione di notizie false e nella persuasione occulta: i messaggi sono stati costruiti usando il linguaggio più adatto ai destinatari e gli stessi destinatari sono stati selezionati a seguito di profilazione e micro-targeting.

https://www.youtube.com/watch?v=AmUC4m6w1wo

"Fake Obama created using AI video tool"

Un falso (ma credibile) Obama simulato tramite una IA dalla University of Washington. BBC News.

(durata: 1 minuto e 26 secondi)

[193] https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9512226.

39.2.1 Il caso Hong Kong e disinformazione social

Tutti abbiamo assistito alle proteste di Hong Kong iniziate il 15 marzo 2019 contro il disegno di legge sull'estradizione di latitanti verso Paesi dove non vi fossero accordi di estradizione. Il timore diffuso riguardava la rottura del precario equilibrio giuridico (noto come "un Paese, due sistemi") tra Hong Kong e la Cina, con il rischio che i residenti di Hong Kong finissero sotto la giurisdizione dei tribunali controllati dal Partito comunista cinese.

Il 19 agosto 2019 Twitter e Facebook hanno denunciato campagne di disinformazione su larga scala attraverso le proprie piattaforme social con immagini alterate e decontestualizzate, con didascalie intese a diffamare e screditare i manifestanti.

Un rapporto dell'Australian international policy institute[194] ha scoperto che la presunta campagna di disinformazione promuoveva tre principali narrazioni fittizie, appositamente ideate per distrarre l'opinione pubblica internazionale, colpevolizzare i manifestanti e assolvere le autorità, invertendo di fatto i rapporti di responsabilità. I 3 filoni narrativi ideati erano: condannare i manifestanti, sostenere la polizia di Hong Kong e insinuare nell'opinione pubblica, soprattutto europea, la tesi del complotto sul coinvolgimento occidentale nelle proteste.

La condanna dei manifestanti avvenne attraverso la disinformazione per trasformare comuni cittadini in haters. Attraverso tecniche di doxxing (diffondere on-line informazioni personali altrui, di solito con intento malevolo), i dati personali, incluse immagini, di giornalisti e di circa 200 manifestanti furono pubblicati online su siti contrari alle proteste (175.000 visualizzazioni uniche). Falsi profili Twitter e Facebook fecero rimbalzare notizie artefatte sui canali social media e i media cinesi diedero ampio risalto a notizie appositamente contraffatte.

Persino i media occidentali e l'opinione pubblica europea si divise fra chi credeva nell'innocenza dei manifestanti e chi li condannava quali terroristi.

39.2.2 Il Caso Cambridge Analytica: come falsare il processo democratico

"Esiste ormai un'industria della persuasione politica che fattura miliardi di dollari: usa potentissimi strumenti informatici e anche psicologici per alterare le scelte dei cittadini. Quando votano, ma non solo. Sono aziende costruite per produrre servizi di propaganda e disinformazione: prendono di mira i singoli individui, ne ricostruiscono idee, abitudini e vulnerabilità attraverso i loro dati personali e li spingono a cambiare comportamento. Facendo, a volte, perfino scelte contrarie ai loro interessi. Cambridge Analytica era all'avanguardia in queste tecniche di manipolazione delle coscienze. Oggi non c'è più, ma esistono altre imprese

[194] https://www.aspi.org.au/report/tweeting-through-great-firewall.

simili" racconta Brittany Kaiser, 33enne texana, gola profonda di Cambridge Analytica, dove ha lavorato per 3 anni e mezzo[195].

"I documenti (relativi alle indagini su Cambridge Analytica) rivelano un'idea molto più chiara di ciò che è effettivamente accaduto nelle elezioni presidenziali degli Stati Uniti del 2016. Esistono prove di esperimenti piuttosto inquietanti sugli elettori americani, di manipolazioni con messaggi basati sulla paura, prendendo come obiettivo i più vulnerabili, e questo sembra continuare" spiega Emma Briant, accademica del Bard College di New York, specializzata in indagini sulla propaganda, che aggiunge che Cambridge Analytica era solo "la punta dell'iceberg".

Relativamente al problema dell'influenzabilità dei risultati elettorali mediante l'utilizzo dei social network, il caso di Cambridge Analytica è particolarmente significativo, dato che è intervenuta in alcuni importanti eventi, tra i quali, oltre alle elezioni federali USA, il referendum sulla Brexit. Al momento però è scientificamente indimostrabile quale sia stato l'effetto realmente ottenuto sui risultati finali.

Cambridge Analytica aveva stipulato un contratto con la società Global Science Research (GSR)[196], in virtù del quale aveva avuto accesso ai dati degli utenti di Facebook (e a quelli dei loro amici) che avevano utilizzato l'app "Thisisyourdigitallife". Ciò aveva permesso alla società di utilizzare i dati di circa 87 milioni di persone per scopi differenti da quelli accademici, per i quali essi erano stati originariamente raccolti. Va aggiunto che i dati erano stati ottenuti partendo da un gruppo di soli 270.000 utenti, ovvero di coloro che avevano usato l'app per effettuare volontariamente il test di personalità ivi proposto.

Dunque, la vera novità introdotta da Cambridge Analytica risiede "nei metodi utilizzati per acquisire i dati personali e nel passaggio dal microtargeting demografico a quello comportamentale che consente campagne molto più efficaci in quanto basate sull'emotività"[197].

Come evidenziato da alcuni studi, "le informazioni ottenute attraverso le risposte ai sondaggi offerti dall'applicazione e i dati sulle preferenze espresse dagli utenti mediante i like hanno consentito di formulare previsioni sugli orientamenti elettorali di questi ultimi sulla base di modelli psicometrici"[198]. Successivamente, "con tali dati e con i risultati dei test, Cambridge Analytica ha potuto applicare i suoi algoritmi di microtargeting psicometrico e di messaggi mirati via social e media tradizionali"[199].

Il Metodo Cambridge Analytica (indicata nel seguito anche come "CA") si componeva di 5 fasi.

---

[195] Brittany Kaiser. *La dittatura dei dati*. Italia: Harper Collins Italia, 2019.
[196] Una parte di questi dati sono poi stato condivisi con il Scl Group, la società madre di Cambridge Analytica che era coinvolta nelle elezioni federali USA.
[197] Colajanni Michele. "Social, raccomandazioni per l'uso". GNOSIS Rivista italiana di intelligence, n. 2/2018.
[198] Alì Antonino. "L'analisi dei dati dei social network per finalità politiche a seguito del caso Facebook-Cambridge Analytica". GNOSIS Rivista italiana di intelligence, n. 1/2019.
[199] Colajanni Michele. "Social, raccomandazioni per l'uso". GNOSIS Rivista italiana di intelligence, n. 2/2018.

1. Segmentazione, attraverso ulteriori elementi:
    - Big data: CA possedeva un database enorme costituito da numerosi database acquisiti online da qualsivoglia fonte (p.e.: Facebook, agenzie creditizie, siti di profilazione, app di terze parti);
    - Igienizzazione dei dati: processo teso a verificare le informazioni e a incrociarle costantemente con quelle nuove;
    - Applicazione del modello Ocean, ovvero di un modello di psicologia comportamentale e sociale ideato dal Dipartimento di psicologia dell'Università di Cambridge; nello specifico le persone vengono segmentate attraverso 32 diversi cluster riconducibili a 5 macrocategorie (O Apertura o Openness, C Coscienziosità o Conscientiousness, E Estroversione o Extraversion, A Disponibilità o Agreeableness, N Nevrosi o Neuroticism) costituenti il profilo della personalità di ogni individuo;
    - Profilazione: una successiva meticolosa attività di profilazione attraverso like sui social media, propensione all'acquisto, punteggio di affidabilità creditizia e altri strumenti affinava ulteriormente la precedente clusterizzazione.

2. Analisi predittiva: Algoritmi predittivi affinavano ulteriormente i dati raccolti arrivando a punteggi di probabilità e accuratezza prossimi al 95%.

3. Intercettazione: Attraverso pubblicità mirate su Twitter, Facebook, YouTube, servizi di streaming musicali e Google, non appena l'utente andava online veniva raggiunto da messaggi (banner, news) appositamente ideati per lui o lei.

4. Intercettazione diretta: Il software Ripon, studiato per la propaganda elettorale casa per casa, consentiva a chi lo utilizzava di accedere ai dati di qualcuno semplicemente avvicinandosi alla sua abitazione o chiamandolo al telefono, approntando una strategia di comunicazione mirata.

5. Microtargeting: l'analisi predittiva assegnava punteggi sulla personalità, gli psicologi determinavano da cosa sono motivate le persone, il gruppo creativo confezionava messaggi specifici veicolati attraverso il microtargeting comportamentale. Ogni singola persona veniva raggiunta da messaggi, video, audio, volantini, tramite un sistema automatico che perfezionava di continuo i contenuti. Statisticamente erano necessarie 20 o 30 diverse versioni dello stesso messaggio, inviate attraverso canali diversi, affinché una singola persona cliccasse e a questo punto CA sapeva esattamente come e dove raggiungere ogni singola persona obiettivo.

Considerando che i sistemi di *(automated) behavioral microtargeting* in futuro "comporteranno conseguenze rilevanti sui meccanismi elettorali e, in ultima analisi, sulla genesi del consenso politico"[200], la tematica è di primaria importanza per la sopravvivenza delle democrazie in un'era iperstorica di post-verità.

Ulteriore elemento di criticità è riferibile al ruolo censorio che i social network, in qualità di società detentrici dei dati in un vero e proprio regime di monopolio, potrebbero esercitare

[200] Alì Antonino. "L'analisi dei dati dei social network per finalità politiche a seguito del caso Facebook-Cambridge Analytica". GNOSIS Rivista italiana di intelligence, n. 1/2019.

mediante meccanismi di “autoregolamentazione” che, essendo esterni all’ordinamento vigente, presentano inevitabilmente profili di opacità[201].

Per un approfondimento a riguardo si raccomanda la consultazione della risoluzione del Parlamento Europeo del 25 ottobre 2018, sull’utilizzo dei dati degli utenti Facebook da parte di Cambridge Analytica e l’impatto sulla protezione dei dati, così come si consiglia la lettura del report dell’Information Commissioner’s Office (ICO) del 11 luglio 2018 “Democracy disrupted? Personal information and political influence”.

[201] Vedere anche la “Risoluzione del Parlamento europeo del 25 ottobre 2018 sull'utilizzo dei dati degli utenti Facebook da parte di Cambridge Analytica e l'impatto sulla protezione dei dati” (https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:52018IP0433) e il report dell’ICO “Democracy disrupted? Personal information and political influence” (https://ico.org.uk/action-weve-taken/investigation-into-data-analytics-for-political-purposes/).

- SESTA PARTE: LE CONTROMISURE

40 La sicurezza dell'IA

40.1 Misure per la sicurezza della logica dell'IA

Per far fronte agli attacchi alla logica dell'IA e all'adversial ML (paragrafo 38.1), negli anni recenti è emerso un campo di ricerca per sviluppare sistemi IA capaci di fornire buone prestazioni anche in ambiente ostile grazie a un apprendimento robusto, capace di far fronte alla presenza di un certo numero di esempi completamente sbagliati da cui imparare e di rinunciare a pronunciarsi se gli esempi che si trova di fronte sono significativamente diversi (a livello statistico) da quelli da cui ha imparato.

Vi è una pressante necessità di metodi formali per verificare le varie componenti dei sistemi di IA, in modo da assicurarne la correttezza logica, anche a fronte delle difficoltà poste dai problemi di indecidibilità e di complessità computazionale.

Per garantire un processo decisionale affidabile è necessario includere metriche sulle prestazioni del processo decisionale stesso.

Si tratta di compiti non di banale soluzione, visto che parliamo di algoritmi che si modificano automaticamente acquisendo informazioni e modelli nuovi anche in autonomia.

40.1.1 Redress by design

Il *redress (riparazione) by design* si riferisce all'idea di stabilire, fin dalla fase di progettazione, meccanismi che garantiscano la ridondanza, sistemi e procedure alternative, ecc. per poter efficacemente rilevare, controllare, correggere le decisioni sbagliate prese da un sistema perfettamente funzionante e, se possibile, migliorarlo.

Come esempio di implementazione del sistema di correzione, si consideri la recente Direttiva UE sul diritto d'autore: un sistema di intelligenza artificiale deciderà se un contenuto è presumibilmente legittimo o se è in violazione del diritto d'autore di qualcuno. Un sistema di filtraggio perfettamente funzionante commetterà degli errori e bloccherà la pubblicazione di alcuni contenuti leciti. I dettagli della procedura di ricorso previsti nella direttiva sono molto scarsi ed è probabile che una normativa ad hoc venga elaborata dagli Stati membri in modo da sfavorire i ricorsi effettivi o la loro efficacia.

Un'alternativa di tipo *redress by design*, ad esempio, avrebbe potuto essere quella di concedere agli utenti la possibilità di opporsi immediatamente a una decisione di blocco e

assicurare la pubblicazione del contenuto a fronte della messa a disposizione di specifiche garanzie valutabili, successivamente, da un tribunale che dovesse riscontrare un illecito.

40.1.2 Ridondanze

Un metodo efficace per scongiurare alcuni attacchi alla logica dell'IA è quello di prevedere una ridondanza del sistema e dei suoi componenti per assicurare controlli incrociati con sistemi paralleli assumendo che chi attacca non riesca a comprometterli tutti.

Questo approccio può applicarsi anche per la difesa dagli *adversarial attacks*: potrebbero essere previsti due algoritmi diversi addestrati con set diversi, ma con livelli di affidabilità simili. In caso di discrepanze tra i due algoritmi, si potrebbe richiedere l'intervento di un essere umano.

40.1.3 Contrasto agli attacchi di "reconnaissance"

Per contrastare gli attacchi di "reconnaissance" (*ricognizione*) con l'obiettivo di apprendere la logica del sistema o di identificarne i punti deboli, si elencano qui di seguito alcuni suggerimenti su misure di prevenzione a questo tipo di attacchi:

- identificare i tentativi di attacco, per esempio dall'eccessivo numero di richieste da una sorgente sospetta;
- a fronte di tentativi di attacco, aumentare il carico di lavoro dell'attaccante in modo da diminuire l'efficacia degli attacchi di reverse engineering del modello;
- controllare gli accessi;
- utilizzare tecniche di honeypotting (creazione di sistemi appositi da far attaccare ai malintenzionati per studiarne le tecniche e ritardarne le attività verso sistemi critici) anche per comprendere gli obiettivi dell'attaccante.

40.1.4 Insiemi dei dati di addestramento

Per ridurre i pregiudizi ed evitare che l'uso dei sistemi di IA porti a risultati che implichino discriminazioni, come anche raccomandato dalla Commissione europea, in fase di addestramento vanno utilizzati dati sufficientemente rappresentativi e qualitativamente adeguati, al fine di garantire che gli stessi rispecchino adeguatamente tutte le pertinenti dimensioni di genere, etnia e altri possibili elementi di possibile discriminazione vietata. Questa soluzione dovrebbe essere richiesta anche dalle future prescrizioni normative in materia di IA.

40.1.5 Prediction poisoning

Il controllo dei risultati di un algoritmo di IA potrebbe prevedere un intervento umano. Ma questo potrebbe essere a sua volta alterato dai risultati dell'IA. Per esempio, in ambito medicale: se un medico dovesse avallare una decisione diagnostica sbagliata di un algoritmo di IA oppure contraddirne una corretta, potrebbe trovarsi in difficoltà nel caso di richieste di chiarimenti. A questo punto ogni medico potrebbe trovare comunque più comodo avallare le decisioni dell'algoritmo in modo da potersi sempre giustificare appoggiandosi ad esso.

Una possibile tecnica di contrasto a questo approccio può essere chiamata "prediction poisoning" (predizioni inquinate) e consiste nel sottoporre all'essere umano, in modo casuale, decisioni sbagliate. In questo modo, l'essere umano sarà sempre costretto a verificare attentamente le proposte dell'algoritmo ed, eventualmente, sottoporle a un terzo parere.

40.1.6 Trasparenza

Infine, si noti come la capacità di verificare l'assenza di *bias* del sistema dipenda anche alla trasparenza delle tecnologie di IA adottate. Tale argomento è oggetto del capitolo 31.

40.2 Sicurezza dell'infrastruttura dell'IA

Come già segnalato nel paragrafo 38.2, gli attacchi possono rivolgersi direttamente ai sistemi che erogano il servizio. Tipicamente si tratta di architetture cloud, quindi diventa fondamentale fornire le protezioni necessarie per mettere in sicurezza l'infrastruttura, le virtual machine o i docker sui quali girano i servizi.

Tra i meccanismi di sicurezza più significativi citiamo:

- firewall e web application firewall; oggi offrono anche verifica del codice javascript e della presenza di eventuali codici bot;
- intrusion detection system; esso può a sua volta far uso di tecniche di intelligenza artificiale imparando a identificare eventuali anomalie e scostamenti rispetto al traffico usuale;
- antimalware;
- hardening;
- controllo degli accessi a più fattori;
- monitoraggio dell'infrastruttura;
- SLA con i fornitori dei servizi cloud e di connettività, se presenti, con tempi ben definiti per la risposta ad eventuali incidenti

La maggior parte dei servizi cloud mette a disposizione le API necessarie per tenere continuamente monitorato il comportamento degli utenti, la tipologia dei dati e il loro utilizzo e fornire l'integrazione con servizi di sicurezza terzi (DLP, antimalware, EDR, ecc).

Da non dimenticare infine la necessità di seguire processi di gestione dei cambiamenti ben definiti e comprensivi di test accurati e completi.

40.3 La continuità dell'IA

Ogni singola applicazione di IA ha requisiti di resilienza e continuità differenti. Un'applicazione per la guida autonoma di un'auto ha esigenze di continuità totalmente diverse rispetto a quella che valuta il merito creditizio di un possibile cliente e la loro indisponibilità comporta analogamente rischi alquanto diversi. L'approccio nella valutazione di tali requisiti non può quindi che partire da una preliminare valutazione degli impatti dell'indisponibilità di tali applicazioni nel tempo, al fine di definire le opportune misure preventive e di ripristino.

Al riguardo vanno evitati gli errori più comuni, derivanti da un uso non critico di metodologie e standard, quali ad esempio quello di valutare un singolo RTO (recovery time objective) e RPO (recovery point objective) mentre questi vanno, ad esempio, legati allo specifico momento (ora del giorno, giorno dell'anno) in cui avviene l'evento che causa l'indisponibilità dell'applicazione.

È anche necessario considerare, insieme all'applicazione IA, anche le sue dipendenze, ossia i processi alimentanti o conseguenti. Ad esempio l'indisponibilità dei primi può portare alla mancanza dei dati che servono all'applicazione di IA per lo svolgimento del suo lavoro.

40.4 Le assicurazioni per l'IA

Nell'ultimo periodo, le aziende utilizzatrici di IA hanno lamentato l'assenza di una formula certa e di *best practices* consolidate che assicurino una mitigazione quasi completa del rischio relativo all'utilizzo dell'intelligenza artificiale, rendendo l'accettazione o il trasferimento del rischio l'unica strada percorribile.

Le cosiddette "cyber insurance" tutelano le aziende e le persone dai rischi relativi all'infrastruttura informatica, alla protezione dei dati personali, alla responsabilità del governo di tali informazioni e alle attività a esse correlate. Tuttavia, le cyber insurance risultano ancora incomplete rispetto ai rischi legati all'IA. Di seguito alcuni esempi di eventi collegati all'IA e potenzialmente coperti da una cyber insurance:

- furto di modelli, ossia la possibilità che qualcuno riesca a ricostruire l'algoritmo di IA partendo dall'output;

- perdita di dati sensibili, come nel caso di incidente relativo al database per il riconoscimento di impronte digitali e dei visi[202].

Alcuni esempi di eventi non coperti, anche considerando le difficoltà di attribuzione della responsabilità civile e penale (capitolo 33):

- danni fisici a soggetti terzi, come nel caso dell'autopilota che, a causa di un malfunzionamento, ha causato un incidente con un pedone (vedere paragrafo 38.1.4.1);
- danni all'immagine, come nel caso del Tay bot di Microsoft (vedere paragrafo 38.1.4.5);
- danni alla proprietà fisica, per esempio un robot pulitore, che utilizza l'intelligenza artificiale per esplorare la casa e, durante il processo, inserisce una scopa bagnata nella presa provocando un corto circuito.

È evidente la necessità di predisporre nuovi prodotti assicurativi, atti a soddisfare le esigenze dei clienti che usano le tecnologie di IA, accompagnati da modelli e controlli appropriati per il calcolo dei premi e dei danni.

[202] https://www.forbes.com/sites/zakdoffman/2019/08/14/new-data-breach-has-exposed-millions-of-fingerprint-and-facial-recognition-records-report/.

41 L’IA per la sicurezza

Il vertiginoso aumento del numero e della complessità degli attacchi registrati negli ultimi anni è divenuto un fattore di crescita determinante per l’utilizzo dell’intelligenza artificiale come strumento fondamentale nella cyber security.

In questo capitolo affrontiamo alcune questioni da considerare quando si usa l’IA per la sicurezza.

Nel prossimo capitolo e nei successivi verrà fornita, quindi, una carrellata sulle aree della sicurezza in cui l’intelligenza artificiale sta già apportando e apporterà sempre di più indiscutibili benefici.

41.1 Il mercato dell’IA per la sicurezza

Il mercato dei prodotti di cybersecurity ha visto negli ultimi anni una corsa all'accaparramento di un qualche tipo di intelligenza artificiale. Ormai ogni vendor ha almeno una soluzione che include il machine learning, vuoi per analizzare nuovi malware, vuoi per trovare nuovi tipi di attacco alle reti. Passeggiare per gli stand agli eventi espone a tutta una serie di dissertazioni su come l'intelligenza artificiale del vendor sia superiore a quella della concorrenza.

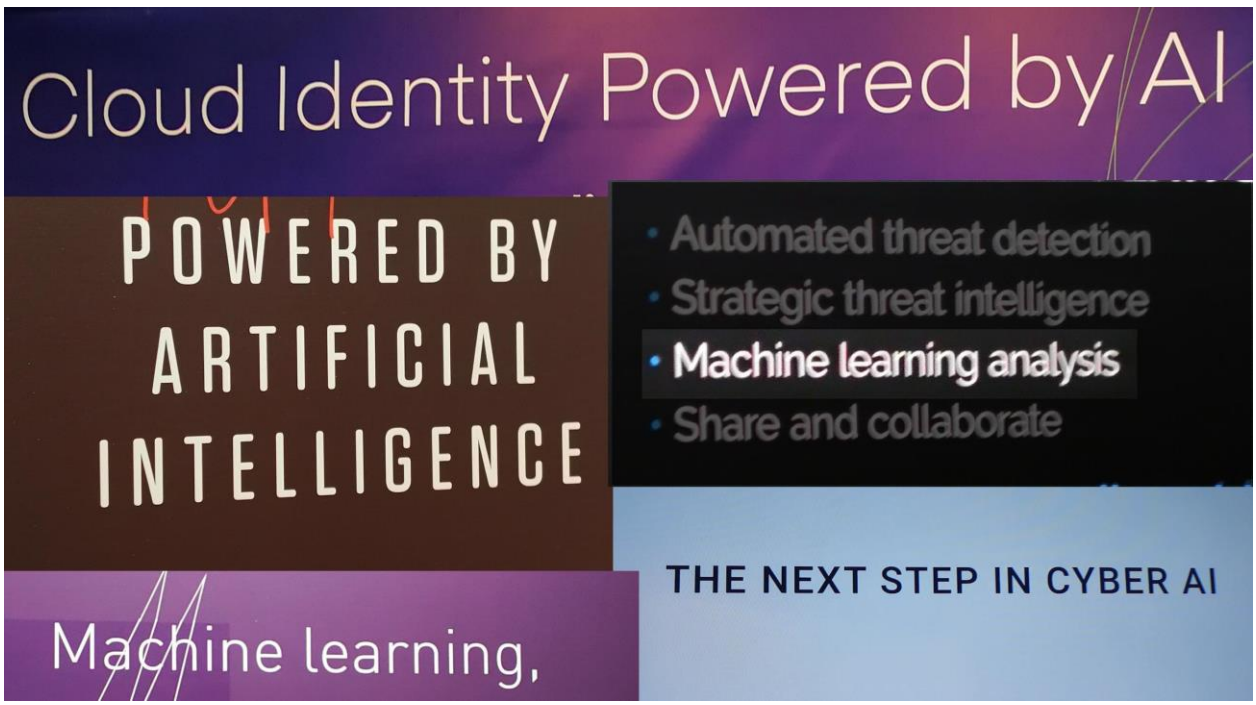


Figura 34 - fotografie scattate durante un evento di cybersecurity a Roma[203]

[203] Foto degli autori.

Non tutti i vendor spingono sul pedale dell'IA con la stessa intensità. Abbiamo quindi deciso di analizzare i messaggi e la documentazione di 40 produttori di software di sicurezza informatica, scelti fra quelli con le migliori valutazioni negli ultimi Magic Quadrant di Gartner e quelli più presenti nel mercato italiano. La nostra ricerca si è basata sulla disamina dei siti web e della documentazione tecnica, inclusi i whitepaper facilmente accessibili al pubblico, simulando un potenziale futuro cliente che cerca informazioni sull'offerta del vendor.

I risultati della nostra indagine mostrano che più di un terzo dei vendor di cybersecurity (il 37,5%) presenta le proprie funzionalità di intelligenza artificiale come forte valore aggiunto e talvolta discriminante della propria offerta. "*La nostra intelligenza artificiale"*, per riassumere il messaggio corale, "*è il motivo per cui dovrai scegliere il nostro software invece che quello della concorrenza"*. Di questi, più della metà (60%) presenta tale fiore all'occhiello direttamente in home page sul proprio sito web per far capire subito ai potenziali clienti che l'IA fa seriamente parte dell'identità del vendor.

Questo in una fase comunicativa fortemente opportunistica: l'indagine infatti è stata condotta nella seconda metà del 2020, quando già molti vendor avevano spostato il loro messaggio principale abbandonando il fulcro tradizionalmente tecnologico ("*la nostra tecnologia è vincente")* per abbracciare quello dell'usabilità e della flessibilità in un'ottica pro-remote working per le conseguenze della Covid-19. Un periodo, insomma, dove si parla meno di intelligenza artificiale e più di *ease-of-use* e adattabilità, specchio dei tempi in un momento dove i compratori sono maggiormente preoccupati dell'improvviso allargamento del perimetro di sicurezza fin dentro le case dei dipendenti.

Eppure molti vendor hanno mantenuto alta l'attenzione sulla propria IA, ed è interessante notare la presenza sia di grandi player come Microsoft e IBM, che hanno investito molto nei loro team di intelligenza artificiale e che quindi applicano le loro tecnologie anche alle business unit che si occupano di security, sia di player più recenti come Darktrace, che usano l'IA come forte differenziatore per emergere in un mercato dove le barriere di entrata sono molto elevate.

Non abbiamo potuto fare a meno di notare, tuttavia, come in diversi casi il messaggio si fermi lì, senza essere ulteriormente raffinato. Un terzo dei vendor infatti - dopo aver sbandierato l'IA come grande valore aggiunto - non fornisce altri particolari. Nel settore dell'intelligenza artificiale si dice scherzosamente che quando si sente parlare di IA generalmente è marketing. Questo perché chi si occupa realmente della tecnologia - che non è altro che un contenitore composto da tante discipline differenti - di solito espone direttamente la particolare tecnica usata. Nel caso della cybersecurity la tecnologia di riferimento è quella del machine learning, che aiuta nell'individuazione di pattern di attacco - come nei firewall e negli intrusion detection system - o nel riconoscere il malware quando gli altri metodi si sono rivelati inefficaci o inconclusivi.

Ebbene, una parte di vendor "virtuosi", per fortuna la maggior parte (due terzi), non si limita a ostentare l'intelligenza artificiale senza fornire dettagli, bensì offre perlomeno un documento esplicativo o un whitepaper per spiegare cos'è la tanto sventolata IA. Lo fa Avast, produttore dell'omonimo antivirus, con una pagina dedicata[204] dove spiega che per individuare nuovo

[204] https://www.avast.com/technology/ai-and-machine-learning.

malware usa delle reti neurali convoluzionali profonde (convolutional neural network o CNN) e un video-commento dal proprio responsabile per l'intelligenza artificiale. Così come lo fa anche Palo Alto Networks, che si affida alle undici pagine di un whitepaper sufficientemente aggiornato[205] per raccontare che i modelli IA usati sono reti neurali semi-supervisionate che, fra le altre cose, imparano a riconoscere i comportamenti dei vari tipi di utenti per minimizzare i falsi positivi.

Infine guardando tutti i vendor presi in esame, che ricordiamo essere i maggiori marchi di cybersecurity sul mercato, stupisce non poco il fatto che quasi la metà di loro non faccia nel modo più assoluto alcun riferimento all'intelligenza artificiale o al machine learning. Fra le ragioni possiamo ipotizzare la volontà di non andare ad alimentare un termine che molti già considerano troppo inflazionato, oltre al pericolo di aumentare la complessità tecnologica di un messaggio marketing che da molte parti si cerca di semplificare. In quei casi parlare di "reti neurali convoluzionali profonde semi-supervisionate" potrebbe avere l'effetto indesiderato di allontanare il compratore anziché avvicinarlo. Per questo motivo diversi vendor, a nostro avviso, avrebbero deciso di evitare troppi riferimenti all'IA, lasciando magari che il concetto fosse solo lievemente accennato dalle tante immagini di cervelli con reti luminose che abbondano ovunque.

In conclusione, poiché l'ondata di intelligenza artificiale ha attraversato il comparto cybersecurity qualche anno prima rispetto ad altri settori, pare che molti vendor di sicurezza abbiano già elaborato una loro maturità nel comunicare il valore apportato dall'IA. Non c'è solo chi spinge sull'acceleratore e fa capire che il prossimo passo della sicurezza è rappresentato senza ombra di dubbio dal machine learning: c'è anche chi accenna alla tecnologia senza però giocarsi tutto e chi preferisce sottrarsi alla corsa, preferendo una comunicazione maggiormente vicina all'utente, più semplice e decisamente meno tecno-soluzionistica.

Intervista a Francesco Bergadano, Professore Ordinario presso il Dipartimento di Informatica dell'Università degli Studi di Torino, docente di Sicurezza Informatica e Direttore del Master Universitario di primo livello in Cybersecurity

Professore ordinario presso il Dipartimento di Informatica dell'Università degli Studi di Torino, direttore del master universitario di primo livello in Cybersecurity.

Domanda 1. Qual è l'impatto dell'IA sulle operazioni anti-cybercrime?

[205] https://www.paloaltonetworks.com/resources/techbriefs/artificial-intelligence-and-machine-learning-in-the-security-operations-center.

L'intelligenza artificiale ha diverse applicazioni nella cybersecurity, che si differenziano a seconda delle tecnologie di IA utilizzate (ad esempio machine learning, natural language processing, agenti intelligenti).

I sistemi di sicurezza che utilizzano machine learning sono strategici per le imprese in quanto sono in grado di prevenire e neutralizzare alcuni attacchi informatici e fronteggiare il cybercrime.

In particolare, gli algoritmi di machine learning consentono di comprendere come gli utenti si comportano e in che modo i dati si evolvono nel tempo. Di conseguenza possono rilevare anomalie (anomaly detection) e generare alert e attivare contromisure automatiche.

Domanda 2. Nei fenomeni di cybercrime, l'anello debole della catena è il fattore umano. Che ruolo può avere l'intelligenza artificiale in questo ambito?

Confermo: ad esempio i ransomware si insinuano nei sistemi sfruttando l'ingenuità dei legittimi utenti dei sistemi.

I ransomware sono programmi informatici dannosi (malevoli) che possono infettare un dispositivo digitale (PC, tablet, smartphone, smart TV), bloccando l'accesso a tutti o ad alcuni dei suoi contenuti (foto, video, file, ecc.), per poi chiedere un riscatto (in inglese, “ransom”) da pagare per “liberarli”.

L’espansione di ransomware e di altri software malevoli può essere contrastata da tecniche di machine learning, che è una tecnologia ormai matura. L'intelligenza artificiale può così consentire la classificazione automatica di contenuti sospetti, che potranno essere cestinati o resi inaccessibile all'utente ordinario.

Domanda 3. Il cybercrime è passato in pochi anni da pochi target di alto profilo, a molti target di basso profilo. Conferma?

I dati sullo sviluppo del cybercrime confermano questo trend.

Gli attacchi a target medio bassi sono spesso meno sofisticati, ma sono distribuiti e diffusi. Di conseguenza tendono a generare una grande quantità di dati, che possono essere analizzati e utilizzati in modo automatico, al fine di ottenere le caratteristiche tipiche di un attacco.

L'intelligenza artificiale può quindi essere sempre più efficace nella protezione dei sistemi IT, sfruttando il potere dell'analisi di grandi quantità di dati per identificare e quindi prevenire questi attacchi.

Domanda 4. Quale suggerimento potrebbe dare alle imprese per utilizzare meglio l'IA nel settore della cybersicurezza?

Il suggerimento che mi sento è di dare alle imprese ha un taglio pratico e tecnico, e ha un duplice profilo:

1. le imprese devono dotarsi e acquistare più di un prodotto di protezione da attacchi che ricorra a sistemi di IA, in modo da non consentire agli avversari che sferrano gli attacchi un preciso vantaggio conoscitivo;
2. le imprese che utilizzano i sopra citati prodotti non devono limitarsi a utilizzare le classiche configurazioni standard ma dovrebbero avvalersi, per questi profili, di professionisti del settore e della possibilità di utilizzare parametri personalizzati, non prevedibili da un avversario.

Domanda 5. Quale è il ruolo dell'università in materia di cybercrime e IA?

L'università anche nel settore dell'intelligenza artificiale svolge tre finalità principali:

1. finalità didattica: l'università insegna i principi e trasmette i metodi e le competenze ai futuri professionisti del settore;
2. finalità di ricerca: avanzamento della conoscenza specifica del settore con finanziamenti europei e nazionali, con scambi e interazioni con università e centri di ricerca di tutto il mondo;
3. l'università dialoga con le imprese e le parti sociali, e individua best practice nella selezione e nel deployment delle tecnologie.

Il Dipartimento di informatica dell'Università degli studi di Torino è stato una delle prime strutture pubbliche di ricerca in Italia a occuparsi di intelligenza artificiale. Rappresenta un'eccellenza dal punto di vista delle pubblicazioni internazionali sul tema, e riporta queste conoscenze nel proprio corso di laurea, con un indirizzo specifico, dedicato al prof. Pietro Torasso, pioniere e punto di riferimento del settore. Organizza inoltre un master universitario in intelligenza artificiale.

Domanda 6. Qual è la situazione del settore dell'IA applicato al cybercrime in Italia?

L'Italia ha eccellenze in questo settore a livello accademico, ma vorrei richiamare l'attenzione sul fatto che non abbiamo, in Italia e in Europa, la presenza di grandi players dell'innovazione come Amazon, Google e Facebook.

Le migliori eccellenze in materia di studio e applicazioni IA, a differenza del passato, non si trovano più soltanto nelle università, ma sempre di più all'interno delle strutture aziendali di questi player, che sono in grado di attrarre i migliori talenti del settore, e muovere grandi capitali e investimenti.

41.2 I limiti dell'IA per la sicurezza

È indubbio che le soluzioni di sicurezza basate su IA (soprattutto ove siano disponibili grandi moli di dati, che permettono di sviluppare un modello di valutazione affidabile), a differenza di quelle tradizionali (che, dobbiamo ricordare, non hanno nessuna funzione predittiva nel caso si verifichi un evento inaspettato) non operano più in maniera deterministica ma introducono il concetto di probabilità, ossia del livello di confidenza relativa alla veridicità della risposta fornita dal sistema di valutazione.

Questo permette di rilevare alcuni limiti.

Il primo riguarda la **comprensibilità delle segnalazioni** (*explainability*). Le anomalie in genere non indicano direttamente quale attacco sia avvenuto, ma semplicemente che un evento è sospetto. Questo problema può essere mitigato attraverso la creazione di modelli specifici di rilevazione delle anomalie il cui output abbia un preciso significato. Inoltre, è possibile unire informazioni su attacchi noti per individuarne varianti. Infine è sempre necessario l'intervento di un operatore addestrato.

Il secondo riguarda l'**accuratezza delle segnalazioni**. Questo introduce il tema del "falso positivo" e del "falso negativo", ossia la possibilità che la segnalazione fatta dal sistema di IA sia non corretta, o che non venga fatta una segnalazione necessaria. Questo fenomeno in molti casi può essere ancora più amplificato in presenza di poche informazioni da elaborare (relative al fenomeno analizzato) e di algoritmi non sufficientemente aggiornati.

Gli algoritmi predittivi di ML rilevano gli scostamenti rispetto a un comportamento abituale (p.e. l'accesso registrato su client o server in orari o giorni insoliti come nel fine settimana, numerosi accessi ravvicinati rispetto al solito e tentativi di attacchi a forza bruta). In base alla loro configurazione segnalano l'evento sospetto con bassi livelli di criticità e di confidenza se basati su pochi dati storici e quindi potrebbe essere un falso negativo. L'accuratezza può essere

inficiata anche perché i dati usati per l'addestramento non sono adeguati o sono stati inquinati da un avversario intelligente.

Viceversa, l'indubbio vantaggio del loro utilizzo è la possibilità di evidenziare situazioni anomale non previste o classificate a priori (si pensi per esempio al tema del "zero day attacks"). È quindi necessario che l'algoritmo fornisca il grado di affidabilità delle sue conclusioni.

Il primo fattore da tenere in considerazione è che, per effettuare un'attività di training degli algoritmi, perché capiscano veramente ciò che è lecito o no (behaviour analysis) in base allo scostamento da operazioni "normali", ci vuole **tempo**, spesso diverse settimane se non addirittura mesi.

Attenzione va posta negli **automatismi** collegati ai sistemi di IA. Se, per esempio, un sistema di IA rileva la necessità di installare un patch e questo viene fatto automaticamente, si potrebbero avere disservizi in orari o giorni molto critici.

Non viene inoltre sottolineato abbastanza che - affinché sia garantita nel tempo l'efficacia di un sistema esperto basato su ML (come potrebbe essere per esempio la rilevazione di una nuova forma di attacco o una nuova tipologia di malware) - è necessario **addestrarlo in maniera continuativa.** Questa attività (piuttosto rilevante nella manutenzione del sistema) è inclusa nei servizi forniti dal produttore, nel caso di un prodotto di mercato, o è a carico dell'utilizzatore, nel caso di una soluzione costruita ad hoc. Nel secondo caso, questa attività non rappresenta ovviamente un limite del ML, ma un costo non sempre inizialmente noto o valutato con la giusta attenzione.

Intervista a Marcello Fausti, Responsabile Cyber Security, Italiaonline

Attualmente responsabile della Cybersecurity del Gruppo Italiaonline.
Nei 10 anni precedenti è stato a capo della Sicurezza ICT di Telecom Italia. Sempre nell'ambito del gruppo, ha maturato esperienze nell'ambito dell'e-business (rapporto con partner, fornitori e canali distributivi) e dei sistemi HR. In precedenza, ha lavorato nel gruppo Olivetti come direttore marketing e business development di una società del gruppo e come marketing manager della divisione PA della capogruppo.

Domanda 1. Quali esigenze in ambito cybersecurity portano a considerare IA e machine learning come un possibile prezioso ausilio?

Il punto da cui vorrei partire è la considerazione che le tecnologie IA/ML (intelligenza artificiale e machine learning), assieme a cloud e big data, sono il motore della rivoluzione digitale che stiamo vivendo. Tutte e tre queste componenti hanno a che fare anche con la cybersecurity; la loro azione congiunta, infatti, sta velocemente producendo un aumento esponenziale della superficie digitale che, come diretta conseguenza, genera un fortissimo incremento dell'esposizione alle minacce.

È il primo postulato della cybersecurity: "più il mondo digitale cresce, meno è sicuro".

Il tratto che più di tutti caratterizza questa dinamica è l'incredibile aumento della velocità del cambiamento. Nuove vulnerabilità, nuovi schemi di attacco, nuovi attori di minaccia, nuove varianti di uno stesso ceppo di virus che disegnano una pandemia digitale contro la quale i vecchi approcci basati sulla capacità di conoscere in anticipo le minacce non è più utile. L'approccio cosiddetto "a firme" è ancora utile per limitare l'ulteriore diffusione di una malattia digitale già nota, ma non è utile a prevenire le malattie che ancora non conosciamo. L'industry della cybersecurity non ha (ancora) messo nel radar un vaccino digitale contro i cyberattacchi ma sta puntando (e forse riuscendo) con approcci innovativi, a costruire strumenti diagnostici raffinati in grado di intercettare fin dal loro primo insorgere le metastasi digitali.

Le vulnerabilità sono un tratto ineliminabile di ogni landscape tecnologico e costituiscono un grande problema. Purtroppo, però, non è sufficiente dare la caccia alle vulnerabilità ma è necessario occuparsi anche delle minacce, dei vettori di attacco, delle tecniche di attacco, insomma, di tutta una serie di piccoli tasselli di un puzzle che, se correttamente ricostruiti, possono dare un'indicazione ragionevolmente fondata dell'insorgere di una minaccia.

Per fare ciò, nella società dei big data, un essere umano (l'analista di cybersecurity) dovrebbe riuscire ad analizzare e correlare velocemente enormi quantità di dati scartando quelli inutili, quelli fuorvianti (sia falsi positivi che falsi negativi) e considerando solo quelli utili a formulare un'ipotesi di (futura) compromissione affidabile. Il nostro analista di cybersecurity, quindi, dovrebbe essere in grado di operare su enormi quantità di dati con accuratezza e velocità e imparando da tutti i falsi positivi e falsi negativi che via via incontra in modo da poter raffinare l'accuratezza delle future analisi. Al fattore di scala a cui è oggi necessario agire, gli esseri umani non possono operare con la necessaria accuratezza e velocità; questo non è un dominio per gli uomini bensì per le macchine.

In casi come questo, l'intelligenza artificiale viene in aiuto tramite il machine learning (molto meglio se unsupervised o deep learning) e il cognitive computing, per operare negli ambiti di incident analysis, threath intelligence e per individuare nuovi IoC (indicator of compromise).

In questo ambito, potremmo sintetizzare la differenza tra l'operato di un essere umano e l'operato di una macchina nel seguente modo: un essere umano esegue un'analisi il cui risultato è ragionevolmente fondato; una macchina, a fronte della stessa analisi, produce un

risultato probabilisticamente fondato. Con tutta la differenza che c'è nell'uso dei due termini sottolineati.

Domanda 2. Quali problematiche stanno emergendo con i primi utilizzi dell'IA nella cybersecurity?

Come conseguenza di questo ragionamento, da qualche anno le aziende che si dedicano alla produzione di soluzioni software per la cybersecurity stanno utilizzando in modo esteso le tecnologie IA/ML che sono molto utili in quanto danno grande visibilità e permettono di fare luce sui primi movimenti di quella che – se non contrastata – diventerà una vera e propria catena di attacco. Ma … ci sono dei ma. Queste tecnologie non sono la panacea per tutti i mali. Se le infrastrutture sono obsolete e vulnerabili non c'è intelligenza artificiale che vi possa aiutare … c'è solo il duro (e a volte durissimo) lavoro di upgrade da fare. Se un attaccante riesce a penetrare le tue difese e a diventare, ad esempio, domain admin, le tecnologie IA/ML possono fare poco.

Tornando alla metafora sanitaria, mi sembra che siamo molto lontani dal poter disporre di un "vaccino digitale universale". In questo momento disponiamo di strumenti che utilizzano un unsupervised machine learning per dare grande visibilità su ciò che accade sulle reti e sugli endpoint, con un'accuratezza che migliora via via che gli algoritmi apprendono cosa è giusto e cosa non lo è rispetto al tuo specifico ambiente.

Insomma, le tecnologie IA/ML possono aiutare molto ma bisogna tenere ben presente che non tutto è gratuito. In molti casi, ci troviamo ad utilizzare tecnologie che producono molti falsi positivi su cui bisogna lavorare. Da questo punto di vista vanno meglio le tecnologie che lavorano lato endpoint rispetto a quelle che lavorano lato infrastruttura di rete, probabilmente perché le prime si basano su ambienti più standard e possono far leva su basi di apprendimento molto ampie (addirittura mondiali). Le tecnologie IA/ML che lavorano sulle reti corporate, invece, soffrono un po' di più e producono parecchi falsi positivi.

Non siamo, quindi, di fronte a tecnologie plug&play ma c'è bisogno di parecchio lavoro per estrarne il valore che, comunque, è indubbio. Possiamo dire che senza questo lavoro a monte, il rischio è che queste tecnologie "perdano una elle" e da plug&PLAY diventino plug&PAY, il che, ovviamente, non è desiderabile.

Domanda 3. In conclusione, quali sono le tue raccomandazioni e quali saranno probabilmente le principali applicazioni future dell'IA?

Un importante effetto dell'introduzione dell'IA nel mondo della cybersecurity è di rendere rapidamente obsoleti i vecchi SIEM, sempre di più relegati al ruolo di semplici gestori di log. I nuovi strumenti di anomaly detection & remediation basati su IA/ML stanno completamente sostituendo i vecchi SIEM nella (forse mai realizzata) pretesa di correlare segnali provenienti da fonti diverse al fine di individuare una potenziale compromissione.

I vecchi SIEM, però, erano aggregatori di allarmi provenienti da fonti diverse. Oggi, invece, rischiamo di trovarci con prodotti molto evoluti che non dialogano tra di loro e che, soprattutto, non sono pensati per integrarsi in console di allarmi pronte per essere gestite esternamente all'azienda (in outsourcing). Se utilizziamo prodotti IA/ML differenti per rete ed endpoint (cosa che, come abbiamo visto in precedenza, ha senso) per analizzare un "caso" potremmo dover usare più di una console.

Insomma, le tecnologie IA/ML applicate alla cybersecurity promettono bene e in alcuni casi mantengono anche le promesse, ma la mia impressione è che siamo ancora in una fase iniziale del loro sviluppo. Mi aspetto che molto ancora sia fatto in termini di efficacia (accuratezza dell'analisi) e di standardizzazione per il dialogo e l'integrazione tra prodotti differenti.

Cionondimeno, utilizzare le tecnologie IA è importante anche perché esse sono utilizzate anche dagli attaccanti. Mentre noi siamo alla ricerca di ipotesi di compromissione da bloccare, loro sono alla ricerca di strade per tentare una possibile compromissione. Le tecnologie alla base sono le stesse.

In conclusione, la transizione digitale è una realtà che porta con sé un nuovo modo di affrontare il business e che ha impatti profondi su chi si occupa di cybersecurity. Controlli di sicurezza in cloud, adozione di tecnologie IA/ML, nuova modalità di gestire il processo di produzione dei servizi digitali e nuovo modo di gestire gli allarmi sono le sfide da affrontare per chi si occupa di cybersecurity.

Chi si occupa di cybersecurity ripone grande speranza nelle tecnologie IA/ML. L'ovvia aspirazione è di avere a disposizione una sorta di vaccino che ci renda immuni da qualsiasi malattia digitale essendo in grado di analizzare e correlare enormi quantità di dati con estrema velocità per scoprire e bloccare istantaneamente le minacce già note o completamente nuove.

La battaglia tra uomini di cybersecurity e attaccanti si sta trasformando in guerra tra macchine dotate di un'intelligenza forse dotata di meno sfaccettature rispetto a quella umana ma certamente capace di elaborare enormi quantità di dati con un grado di accuratezza crescente e con una enorme velocità. È questo il nostro futuro?

42 L’IA nei controlli di sicurezza informatici

42.1 Identificazione degli eventi malevoli

Come confermato dagli analisti di settore, il traffico Internet legato al business delle aziende aumenterà di tre volte nel lasso temporale 2017-2023. Sarà quindi sempre più difficile per gli analisti di sicurezza poter monitorare efficacemente i volumi, le velocità e le varietà di dati e log generati dai presidi di sicurezza aziendali (p.e. firewall, IDS/IPS, SIEM) e le informazioni provenienti dall’esterno (p.e. indicatori di compromissione forniti da fornitori di servizi di sicurezza, condivisi sui canali di information sharing o su piattaforme di cyber threat intelligence).

Le soluzioni basate su IA sono, purtroppo, ancora all’inizio degli sviluppi e pertanto vanno sempre affiancate a soluzioni tradizionali.

Alcuni ricercatori stimano che, grazie a questo affiancamento, si possa arrivare a tassi fino a quasi il 100% della rilevazione degli eventi malevoli con un minimo di falsi positivi. L’IA riesce infatti a filtrare le informazioni che non hanno particolare rilevanza per lo specifico contesto (potrebbero invece averla per altre organizzazioni) riducendo i falsi positivi e focalizzando l'analisi sulle informazioni realmente utili.

L’IA nei servizi SOC

L’obiettivo è l’automazione di processi e tecnologie di sicurezza per consentire agli operatori e analisti di concentrare l’attenzione solo su aspetti che non possono prescindere dall’intervento umano.

Le tecnologie sono sempre più tra loro integrate (si parla di SOAR, *security orchestration, automation and response*). Strumenti di sicurezza non integrati richiedono una molteplicità di conoscenze specialistiche, anche relative alle specifiche piattaforme, non sempre disponibili da cui consegue una maggiore difficoltà nel comprendere la natura di un evento. Le tecnologie integrate permettono invece di gestire in maniera automatizzata il ciclo di vita di un incidente (rilevamento, gestione, analisi, risposta, valutazione dell’intervento) con una riduzione sostanziale dei tempi d’intervento anche grazie alla capacità di automatizzare e velocizzare i task ripetitivi. Esse permettono, ad esempio, di arricchire (data enrichment) le informazioni grezze raccolte con gli strumenti di rilevazione (detection), applicare tecniche di analisi forense, analizzare il malware con il supporto di sandbox, produrre documentazione automatica e classificare la gravità della minaccia o dell’attacco.

Le tecnologie di supporto all’identificazione degli incidenti, in questi ultimi anni, sfruttano anche algoritmi di machine learning, in grado di fornire, attraverso un continuo apprendimento sullo storico degli incidenti e sui loro attributi. Le tecniche tradizionali, basate su *signatures* o IoC

(indirizzi IP, email, domini, hash, ecc.), sono molto efficaci sulle minacce già note, ma risultano meno efficaci nel contesto attuale.

L'IA può essere sfruttata in più modi riportati nel seguito.

1. L'IA combina dati (strutturati) provenienti dai sistemi dell'organizzazione con i dati (anche non strutturati) provenienti da fonti aperte come siti web, blog, social network, forum, chat, repository storici di attacchi e vulnerabilità e altre fonti anche solo parzialmente accessibili. Questo utilizzo dell'IA consente di portare alla luce informazioni che risultano predittive di tendenze o di fenomeni di sicurezza che stanno caratterizzando gli scenari tecnologici del momento e che possono avere un impatto sull'organizzazione.
2. Grazie all'aumento dei log, l'IA crea modelli comportamentali dei sistemi, degli utenti e degli amministratori dei sistemi. Le soluzioni UEBA (*user and entity behavior analytics*), solitamente integrate con i SIEM, sfruttano algoritmi di machine learning per correlare le interazioni tra utenti, sistemi, applicazioni, IP e dati al fine di creare delle baseline e classi (o cluster) di comportamenti e allertare in caso di scostamenti (p.e. traffico anomalo) che potrebbero essere originati da attacchi provenienti dall'esterno, botnet sulla rete, malware o comportamenti anomali degli utenti.
3. Il machine learning e l'IA sono impiegati a bordo degli endpoint e analizzano il comportamento dei file eseguiti, le pagine web e il codice eseguito in memoria.
4. L'IA comprende i modelli di minaccia già rilevati per identificare nuovi attacchi e, quindi, ridurre il tempo e gli sforzi per l'identificazione di incidenti e per le successive fasi di risposta, rimedio e indagine.

Le tecnologie di IA permettono quindi di svolgere numerose attività di supporto all'identificazione e gestione degli incidenti elencate nel seguito.

1. Riconoscere, con attività di cyber threat intelligence e di analisi comportamentale, in maniera tempestiva eventi malevoli, ossia nuovi malware o tentativi di attacco, che, per la loro novità o complessità tecnica, non è semplice identificare con le metodologie più tradizionali.
2. Correlare le informazioni raccolte in modo da stimare, in maniera non solo automatica ma anche dinamica, le possibili conseguenze di una minaccia o di un attacco e, quindi, classificarli e assegnare loro una priorità.
3. Indicare le migliori azioni e procedure che possono essere applicate per risolvere il caso; questo anche perché possono essere evidenziate similitudini tra incidenti diversi che possono essere gestiti sfruttando approcci e tecniche risolutive già utilizzate e andate a buon fine.
4. Assegnare le attività di analisi agli operatori più appropriati per competenze e per esperienze pregresse maturate su casi similari.
5. Alimentare continuamente cruscotti direzionali per visualizzare un quadro sintetico sul livello di sicurezza dell'organizzazione e cruscotti operativi per fornire dettagli tecnici e quindi migliorare la capacità di prevenire e reagire agli attacchi informatici, analizzare scenari anomali e visualizzare indicatori di prestazione.

Ai fini della sicurezza informatica e delle frodi, una sfida particolarmente interessante è la gestione dei dati grezzi in modalità streaming, ossia generati ad alta velocità e da un gran numero di sorgenti eterogenee, strutturate e non, per effettuare analisi in tempo reale.

42.2 Identificazione delle vulnerabilità

Grazie alle capacità di auto-apprendimento, l'IA è in grado di simulare attacchi sofisticati in modo automatico e continuativo, modificando e adattando le proprie azioni in base allo specifico contesto (p.e. infrastruttura, capacità di difesa) proprio come fanno gli attaccanti. Alcune soluzioni prevedono, ad esempio, l'utilizzo di sonde in grado di apprendere autonomamente il comportamento della rete e degli applicativi al fine di identificare le vulnerabilità e sferrare gli attacchi in modo efficace.

I sistemi basati su IA (più precisamente reti neurali, algoritmi di ML e in particolare di deep learning), sono in grado, oltre a svolgere le attività di rilevazione delle vulnerabilità note, di identificare vulnerabilità zero-day attraverso l'analisi di anomalie e comportamenti sospetti nei sistemi; identificazione difficile da attuare con sistemi tradizionali di VA che si basano generalmente su firme note e analisi statica o dinamica dei malware.

Alcuni strumenti valutano in maniera intelligente anche il livello di rischio associato a vulnerabilità riscontrate in una determinata infrastruttura, non solo in base al tradizionale CVSS, ma anche alla configurazione della rete stessa, alla probabilità di realizzazione effettiva dell'attacco, ecc.

42.3 Correlazione e risposta agli eventi malevoli

Come visto nei paragrafi precedenti, i sistemi di intelligenza artificiale permettono di identificare gli attacchi informatici. Essi possono guidare anche la risposta agli attacchi, per esempio attraverso funzionalità di self-patching o auto-riconfigurazioni automatiche.

Algoritmi basati sul supervised machine learning vengono utilizzati per attivare meccanismi di dynamic application containment (blocco dell'applicazione o tracciamento delle attività) sulla base di un punteggio o della reputazione dell'applicazione stessa.

Le analisi comportamentali permettono di contrastare:

- i malware polimorfici che stanno rendendo sempre più inefficaci i tradizionali sistemi antimalware basati sulle firme;
- gli attacchi fileless, quindi eseguiti prevalentemente in memoria RAM;
- il rilevamento delle varianti di attacchi già noti attraverso anche il pre-execution machine learning (analisi del codice che permette di bloccare i file malevoli prima della loro esecuzione).

Grazie all'automatizzazione delle attività di routine, il personale addetto alla sicurezza può quindi dedicarsi allo sviluppo di opportune strategie che proattivamente mettano in atto le più adeguate misure di sicurezza in risposta alle minacce esterne ed interne.

Le soluzioni integrate più avanzate fanno leva su *playbook*, che consentono ai team della sicurezza di plasmare e adattare automaticamente i processi di risposta agli incidenti alle peculiarità dell'organizzazione. L'IA e le tecniche di autoapprendimento riescono a gestire *playbook* aggiornabili dinamicamente che sono in grado di estrapolare, in alcuni casi anche graficamente, le relazioni tra IoC (indicator of compromise) e incidenti già noti. Questo permette agli analisti di agire sulle priorità più alte, visualizzando anche tutti gli eventi appartenenti allo stesso incidente.

In questo modo si riesce a ridurre i tempi di risoluzione degli incidenti a pochi secondi o minuti, rispetto alle settimane o mesi richieste nel recente passato.

42.4 Identity and access management

Come ha dichiarato Forrester Research[206], la risposta delle imprese a violazioni di dati, intrusioni malevole e frodi richiederà il ricorso a soluzioni di identity management in grado di rilevare dinamicamente eventuali attività anomale.

Lo scenario è diventato ancora più complesso durante il recente periodo della pandemia da COVID-19, con l'aumento del lavoro da remoto e dell'uso del cloud, con accessi da molteplici punti e dispositivi, spesso non gestiti dall'organizzazione. Si è anche resa necessaria una maggiore apertura di applicazioni e reti a terze parti e clienti.

Per questo motivo, le tecnologie IA e ML sono sempre più spesso adottate nell'ambito IAM (identity e access management)[207]: in aggiunta alla biometria, al controllo avanzato dell'identità degli utenti (basato su riconoscimento di suoni, immagini, comportamenti specifici dei singoli), IA e ML offrono il vantaggio di automatizzare attività altrimenti di competenza di specialisti umani. Nell'ambito della gestione degli accessi e delle identità, il numero degli eventi da considerare può infatti essere elevatissimo.

Tra i vantaggi nell'uso dell'IA per l'IAM:
- il maggior livello di automazione e di flessibilità;
- la possibilità di approfondire ogni singolo dettaglio di una richiesta di accesso: giorno e ora, tipologia di dispositivo, sede, risorse a cui si vuole accedere;
- maggiore visibilità di eventi significativi (una macchina può identificare e analizzare numerosissimi eventi, in modo granulare, in qualsiasi momento e con grande velocità) in modo da inviare un avviso in caso di comportamenti anomali, lasciando che siano poi le persone fisiche a prendere le decisioni.

Sono in uso sistemi di intelligenza artificiale applicati a:

[206] Andras Cser, Sean Ryan with Merritt Maxim, Benjamin Corey, Peggy Dostie. The Top Trends Shaping IAM In 2020. Forrester, 2020.
[207] Artificial Intelligence and Machine Learning are Transforming IAM, Identity Management Institute, https://www.identitymanagementinstitute.org/artificial-intelligence-and-machine-learning-are-transforming-iam/

- **identificazione degli utenti**; gli algoritmi di ML possono supportare la verifica dei documenti d'identità nella fase di registrazione di un utente e ridurre il rischio di furto di identità digitale;
- **autenticazione con dati biometrici**, approfondita nel seguito;
- **autenticazione adattativa**, cioè alla contestualizzazione dell'autenticazione allo scopo di evidenziare gli indicatori di rischio e, conseguentemente, richiedere, se necessario, all'utente ulteriori elementi di autenticazione (risk-based authentication);
- **entitlement management**, che automatizza i processi di richiesta di accesso, attribuzione dello stesso, riesame e scadenza delle credenziali, assegnazione dei ruoli;
- **rilevazione di anomalie** degli accessi, a seguito dell'analisi del comportamento degli utenti (utilizzando informazioni come ad esempio orari di accesso, tipi di dispositivi usati, localizzazioni geografiche, siti acceduti e anagrafica).

42.4.1 Predictive identity

Si comincia oggi a parlare di soluzioni di predictive identity, basate su IA e ML e in grado di predire trend e comportamenti ricorrenti. Le soluzioni IAM tradizionalmente hanno fatto sempre uso di analytics (basati su algoritmi evoluti di data mining) il cui fine era quello di creare un modello di regole standard e permettere di assegnare le opportune autorizzazioni alle persone sulla base dei rispettivi ruoli. Tali sistemi, oltre a richiedere competenze avanzate, non hanno avuto grande successo a causa dell'enorme mole di attività e analisi manuali richieste[208].

Grazie all'adozione di nuove tecnologie che fanno leva anche su IA e ML, la predictive identity promette ora di produrre in automatico modelli di gestione degli accessi digitali pronti per essere riesaminati e validati e di automatizzare l'assegnazione delle autorizzazioni alle identità digitali in occasione di assunzione, cambi organizzativi e uscita dall'organizzazione.

42.4.2 Autenticazione con dati biometrici

L'autenticazione di un utente con dati biometrici può essere sostanzialmente di 5 tipi:

- riconoscimento dell'impronta digitale;
- lettura dell'iride;
- riconoscimento facciale;
- riconoscimento del timbro della voce;
- analisi comportamentale.

[208] Paolo Tarsitano. Predictive Identity: soluzioni di sicurezza per la gestione delle identità digitali e degli accessi. 30 Giugno 2020. https://www.cybersecurity360.it/soluzioni-aziendali/predictive-identity-soluzioni-di-sicurezza-per-la-gestione-delle-identita-digitali-e-degli-accessi/

In generale si tratta di meccanismi più sicuri di quelli basati su password o oggetti fisici, anche se non tutte le soluzioni sono equivalenti. Ad esempio, esistono tecniche per fotografare le impronte digitali lasciate sul telefono della vittima e utilizzarle sul lettore di impronte per ottenere l'accesso. Da questo punto di vista il riconoscimento facciale di tipo tridimensionale è più sicuro, anche se si è dimostrato che può essere ingannato, ad esempio usando stampe 3D basate sulla fotografia 2D del soggetto o, utilizzando la somiglianza fisica.

Alcune di queste tecniche fanno in maniera esplicita ricorso all'intelligenza artificiale. Si pensi ad esempio al riconoscimento del timbro della voce, che è un parametro personale come una impronta digitale. È anche possibile integrare dati biometrici multipli: ad esempio riconoscimento facciale associato a un'impronta digitale o al riconoscimento della retina. In questo caso la sicurezza risulta notevolmente superiore.

Alcuni dati possono aiutare a comprenderne meglio l'importanza di quanto detto. A oggi si stima che quasi il 60% delle app disponibili includa un'opzione di login di tipo biometrico. Quella di gran lunga preferita dagli utenti (oltre il 60%) è il riconoscimento dell'impronta digitale, perché più comoda da usare, ma, come detto, non necessariamente la più sicura. Circa il 50% di utenti americani usa il riconoscimento biometrico per effettuare pagamenti e un altro 40% non utilizzerebbe un'app bancaria che non preveda accesso biometrico[209]. A questo proposito ricordiamo che l'alternativa con SMS non è altrettanto sicura, e che questo tipo di autenticazione può essere più facilmente violata di una basata sul riconoscimento biometrico.

42.5 Compliance agli standard e alle normative

Le tecniche di machine learning, come dimostrato da alcune esperienze significative[210], possono essere di grande utilità nel garantire la sicurezza dei sistemi dove venga richiesta la conformità a qualche normativa o a standard di sicurezza (per esempio in ambito finanziario o sanitario).

Questi strumenti sono in grado di riconoscere automaticamente i dati di interesse (ad esempio, dati particolari come quelli sanitari) e di fornire all'organizzazione cruscotti, sistemi di allarme e piena visibilità su come e quando il dato è acceduto, modificato o spostato. Il monitoraggio continuo delle attività che riguardano i dati genera allarmi dettagliati in caso di anomalie che potrebbero indicare un tentativo di accesso non autorizzato o di una violazione.

[209] https://www.computersciencezone.org/biometric-security/.
[210] https://www.finextra.com/blogposting/16050/ai-and-ml-in-financial-services-compliance-management-use-cases-for-fis.

42.6 L'IA a supporto della continuità

L'IA può essere utilizzata al fine di garantire la continuità in caso di guasto, disastro o perdita di dati.

L'IA è in grado di analizzare i dati e prendere decisioni con una rapidità di gran lunga maggiore rispetto anche al professionista umano più abile. Per questo si prevede che potrà essere usata per automatizzare la continuità operativa di un'organizzazione.

Va detto, tuttavia, che non tutti i processi possono essere automatizzati. In particolare, quando si verifica una crisi, devono essere prese decisioni repentine e complesse; decisioni che richiedono una combinazione, che l'IA non ha finora raggiunto, di intelligenza logica ed emotiva[211].

Una delle applicazioni già attualmente in uso dell'IA è quella di rendere più efficaci i cicli di backup perché non più legati a programmazioni statiche[212].

42.7 L'IA a protezione delle email

Le piattaforme di scansione multipla e simultanea delle email sono in grado di individuare e rimuovere malware provenienti da più fronti, facendo uso di intelligenza artificiale e machine learning per localizzare la minaccia, intercettare tentativi di accesso non autorizzati, analizzare i file sospetti, scomponendo il file, rimuovendo parti potenzialmente pericolose e ricomponendo il contenuto in un file rigenerato.

Tali piattaforme permettono anche di identificare email di spam o fasulle come BEC (business email compromise), CEO fraud, messaggi sui social da account finti e spear phishing.

L'IA può aiutare nella classificazione dei contenuti e si sta avvicinando (grazie al natural language processing o NLP) a una comprensione reale del testo dell'email, quindi ad afferrarne il senso e a identificare aspetti come: il carattere di urgenza, autorità presunta da cui deriva la richiesta e il tipo di richiesta, eventuale richiesta di non essere ricontattati. Tutte queste informazioni vanno ad aggiungersi alle tradizionali informazioni analizzate dal filtro antispam, permettendo di avere così un più elevato livello di identificazione nei confronti di BEC, di phishing o di truffe in genere. Oltre all'analisi del testo, sfrutta anche l'analisi delle pagine web riportate negli eventuali link contenuti nelle mail.

Le soluzioni di IA sono anche in grado di creare un modello dello stile di scrittura, per riconoscere se una mail è stata realmente scritta dal mittente indicato nella stessa. Questa

[211] https://www.riskmanagement360.it/enterprise-risk-management/il-ruolo-dellai-in-business-continuity-e-risk-management/.
[212] https://aibusiness.com/author.asp?section_id=789&doc_id=761245.

tecnica è utilizzabile soltanto per i mittenti facenti parte dell'organizzazione che utilizza le piattaforme di analisi, poiché l'IA ha bisogno di accedere alla cassetta postale del mittente per poterne analizzare lo stile di scrittura. Purtroppo tali tecniche non sono ancora completamente adattate per la lingua italiana.

Le soluzioni tradizionali prevedono di applicare:

- ai parametri dell'header (la parte dell'email che contiene informazioni tecniche, come il mittente e l'IP di provenienza) regole basate su liste di controllo (ad esempio IP malevoli o provenienza non abituale) e implementando sui server di posta di controlli anti-spoofing (SPF, DKIM, DMARC) con criteri di hard fail e rifiuto ove applicabili, per verificare la veridicità di quanto contenuto in una mail, benigna o malevola[213];
- al body (in cui è presente il testo vero e proprio del messaggio) tecniche per cercare di interpretarne il contenuto e ricondurlo a schemi che consentano di attribuirgli un grado di rischio relativo al phishing.

La combinazione delle analisi dell'header e del body permettono di classificare un'email come genuina, phishing o sospetta.

Per analizzare il contenuto del body della email furono inizialmente applicati algoritmi per il riconoscimento semantico del testo. A partire dal 2014, però, si affermarono algoritmi che sfruttano i principi delle reti neurali e che producono un minor numero di falsi positivi. Alcuni produttori adottarono quindi reti neurali convoluzionali (convolutional neural networks o CNN), da tempo già utilizzate per il riconoscimento di immagini, applicandole al natural language processing (NLP) e creando delle componenti di riconoscimento del phishing, spear phishing e business email compromise. Questo senza far uso di liste di controllo o regole statiche.

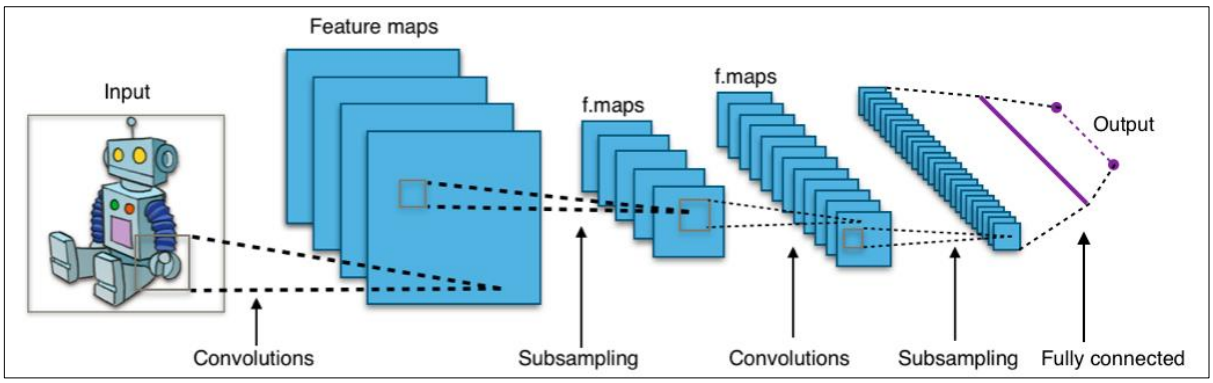


Figura 35 - Rete neurale convoluzionale[214]

Basandosi su un modello R-CNN (recurrent convolutional neural network) con vettori multilivello e meccanismo di attenzione, è possibile realizzare un sistema di rilevamento del phishing che utilizza contemporaneamente l'header e il body del messaggio sia a livello di

[213] https://www.garrnews.it/cybersecurity-22/cybersecurity-month/820-phishing-nell-emergenza-covid-strategie-di-difesa-aziendale
[214] https://it.wikipedia.org/wiki/Rete_neurale_convoluzionale.

carattere sia di parola[215]. Questo permetterebbe di raggiungere una percentuale di successo nell'individuazione del phishing molto elevata e che, nei test di laboratorio, si è attestata intorno al 99%.

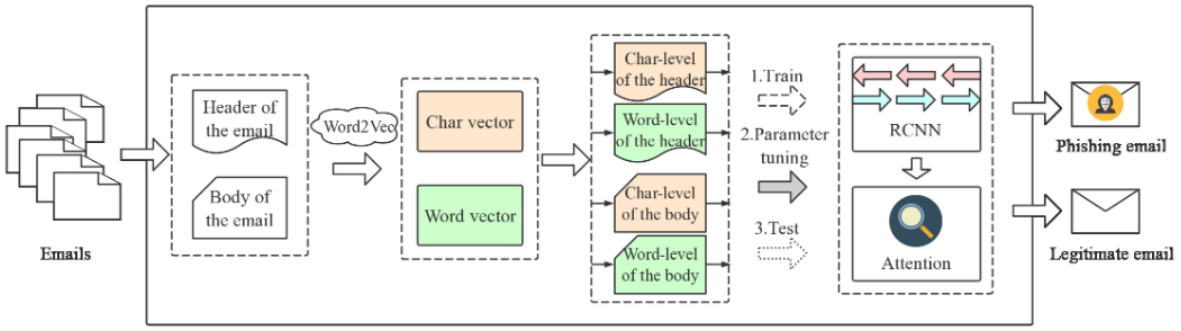


Figura 36 - Phishing email detection using improved RCNN model[216]

42.8 Sicurezza delle comunicazioni

Analisi dei nomi di dominio

Un ambito di applicazione dell'intelligenza artificiale è rappresentato dall'analisi dei nomi di dominio contattati dagli host. Essa può essere utile sia per la rilevazione dei malware (nomi di dominio ad hoc vengono utilizzati dalle botnet per nascondere i server di command & control) sia per la rilevazione di attacchi contro gli utenti, siano questi mirati (come gli APT) oppure campagne di phishing su larga scala.

In questo caso, trovano sicuramente applicazione in ambienti reali tecniche di analisi passiva del traffico DNS (ossia tecniche di raccolta del traffico prodotto dagli utenti verso i recursive DNS) che prevedono l'utilizzo di algoritmi di clustering al fine di creare blacklist di domini malevoli.

Web content filtering

Un'ulteriore applicazione dell'IA è quella del web content filtering, con cui è possibile effettuare su Internet ricerche di documenti specifici o filtrati per categoria (p.e. "minori"), facendo uso di motori di ricerca sperimentali, specifici per argomento. In questo modo è possibile, tramite reti neurali e macchine vettoriali, per mezzo di algoritmi di apprendimento automatico e con un approccio basato sul lessico (p.e. dizionario di vocaboli), analizzare e filtrare i documenti più rilevanti da documenti irrilevanti o da oscurare (p.e. contenuti offensivi, violenti o per adulti).

---

[215] Yong Fang , Cheng Zhang, Cheng Huang , Liang Liu, Yue Yang. Phishing Email Detection Using Improved RCNN Model With Multilevel Vectors and Attention Mechanism. *IEEE Access*. 2019. doi: 0.1109/ACCESS.2019.2913705.
[216] Dall'articolo già citato.

IA e web content filtering potrebbero in futuro essere usati come strumento intelligente per risolvere le disinformazione che corre in rete a causa di fake news e bufale sui social network.

42.9 L'IA e la sicurezza delle applicazioni software

Con l'espressione *sicurezza applicativa* si fa riferimento a un ampio ventaglio di soluzioni, metodologie e pratiche che hanno il fine di garantire la sicurezza delle applicazioni e dei dati in esse ospitati. La sicurezza applicativa mira a garantire la continuità di funzionamento dell'applicazione (disponibilità), proteggere i dati da essa gestiti, garantendo il soddisfacimento dei requisiti di riservatezza e integrità, mantenere l'integrità del codice dell'applicazione.

Eventuali vulnerabilità o debolezze dell'applicazione che l'attaccante può sfruttare a proprio vantaggio per compromettere l'applicazione possono essere individuate e risolte in diverse fasi del ciclo di sviluppo e di utilizzo di un'applicazione, anche con il supporto di algoritmi di intelligenza artificiale o di machine learning.

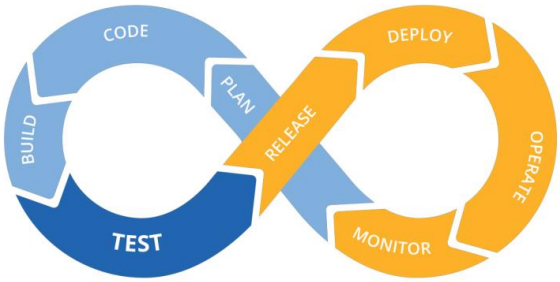


Figura 37 - Il ciclo di vita del software secondo il modello DevOps[217]

Code review automatico con IA

Nella fase di sviluppo, è pratica comune avvalersi di soluzioni per la ricerca di vulnerabilità all'interno del codice sorgente. Come esempio possiamo portare l'iniziativa e l'esperienza della fintech DeepCode che ha creato un servizio erogato in cloud, basato sul machine learning, che analizza il codice C e C++ alla ricerca di potenziali difetti e bug.

Addestrato attraverso l'analisi di centinaia di progetti open source, ha dimostrato la sua superiorità rispetto ai tradizionali strumenti di analisi del codice perché viene eseguita una analisi di contesto, cioè relativa all'esecuzione del codice e non alla sola analisi testuale del codice stesso.

In-app protection: analisi comportamentale per la sicurezza dell'app

[217] Fonte: The Definitive Guide to DevOps, Project-Management.com 2019 (https://project-management.com/the-definitive-guide-to-devops/.

Nell'ambito delle applicazioni ci si pone l'obiettivo di renderle immuni da malware o attacchi durante la loro esecuzione; questo sia per evitare la compromissione dei dati personali dell'utilizzatore sia per il contrasto a possibili frodi.

Un'app deve essere, per esempio, in grado di bloccare in maniera autonoma tentativi di infrazione (funzione nota come RASP, runtime application self-protection). Questo avviene con il supporto di un servizio cloud che raccoglie le informazioni fornite dall'app stessa, le elabora in tempo reale, mantiene una fotografia dell'ambiente in cui l'app gira e verifica, mediante tecniche di analisi comportamentale, che l'utilizzatore dell'app sia quello abituale e non un malware o un'altra applicazione malevola.

Queste analisi oggi sono fatte grazie all'utilizzo di algoritmi di machine learning che, dopo un breve periodo di training, sono in grado di rilevare le anomalie e gli scostamenti rispetto al profilo standard, fornire una gravità di quanto rilevato e, ovviamente, indicare le eventuali contromisure da adottare per bloccare o mitigare il problema.

Protezione delle applicazioni con web application firewall e IA

Una volta passati alla fase di esercizio, le applicazioni web e le API vengono spesso protette da un web application firewall (WAF).

Nei moderni WAF, l'intelligenza artificiale e il machine learning sono utilizzati per costruire, a partire dal traffico in ingresso ai servizi web, un modello di funzionamento legittimo dell'applicazione, ossia un modello di come l'applicazione dovrebbe comportarsi a fronte di un normale e corretto utilizzo da parte degli utenti (ad esempio, stabilire quali sono i valori tipicamente ricevuti da un particolare parametro). Tale modello viene applicato al traffico in ingresso, consentendo di identificare i tentativi di attacco rappresentati da richieste anomale rispetto al modello creato.

Analisi predittive con l'IA sulle vulnerabilità del software

L'approccio basato sull'utilizzo di algoritmi di machine learning può essere efficacemente applicato anche alla gestione delle vulnerabilità di tipo applicativo.

Le organizzazioni lavorano in un contesto caratterizzato dalla segnalazione di circa 12.000 nuove vulnerabilità all'anno e si ritrovano quindi a fronteggiare i seguenti problemi:

- la necessità di allocare risorse finite a un numero elevato di patch da installare;
- attacchi portati sfruttando anche vulnerabilità classificate a livello medio di gravità;
- il bisogno di assegnare le opportune priorità alle attività relative al piano di rimedio;
- la relativamente lunga finestra temporale necessaria alle attività di rimedio in cui risulta possibile a un attaccante realizzare l'exploit funzionale a sfruttare la vulnerabilità.

Utilizzando algoritmi di machine learning impiegati da alcune soluzioni già in commercio, è possibile spostare la prospettiva dal punto di vista dell'attaccante: si analizza in dettaglio la base dati disponibile sulle vulnerabilità e la si incrocia con la disponibilità di exploit disponibili in rete tramite analisi di threat intelligence e rilevamento di attività e conversazioni ostili sui social media.

42.10 L'IA a protezione dell'IoT

Gli oggetti su Internet eseguono compiti ben chiari e definiti, ma potrebbero essere aggiornati con funzionalità non previste. Per esempio, un normale sensore di temperatura di un acquario (connesso a una rete wi-fi) può consentire l'accesso alla rete da parte di malintenzionati[218]; se il sensore continua a dare le informazioni di temperatura dell'acquario, nessuno si accorge dell'anomalia.

Figura 38 - The ransomware of things[219]

Un monitoraggio delle attività degli oggetti IoT grazie l'IA permette di costruire un modello di funzionamento che identifichi il comportamento corretto dell'oggetto (anche considerando i consumi energetici[220]) che può variare nel tempo. L'algoritmo deve evolvere in modo tale da non dare falsi allarmi quando un aggiornamento del dispositivo rende attuali nuove funzionalità del tutto benevole.

[218] https://www.washingtonpost.com/news/innovations/wp/2017/07/21/how-a-fish-tank-helped-hack-a-casino/
[219] https://marketblog.me/2017/05/30/my-take-on-the-ransomware-of-things/.
[220] https://www.aimsciences.org/article/doi/10.3934/mfc.2018004

Quando invece l'oggetto comincia ad avere comportamenti sospetti, non usuali e potenzialmente malevoli, il monitoraggio deve attivare un sistema di allarmistica e un sistema di contenimento dell'attività dell'oggetto (per esempio mettendo gli oggetti connessi su reti dedicate e separate).

Ulteriori considerazioni riguardano i dispositivi mobili (smartphone e tablet) usati per controllare gli oggetti. Anche per questi ML e IA possono essere utilizzati per rilevare il malware[221].

[221] https://dl.acm.org/doi/10.1145/2818000.2818038.

43 L'IA per la prevenzione delle frodi

Il Codice civile non definisce in modo esplicito la frode, limitandosi a configurare solo l'ipotesi di contratto in frode alla legge, fattispecie di illiceità nella causa di un contratto, contraria a norme imperative (art. 1344).

Il Codice penale è più preciso, in quanto, pur non parlando esplicitamente di frode, qualifica la fattispecie di truffa, ossia quando uno o più soggetti, con artifizi o raggiri, inducendo taluno in errore, procura o procurano a sé o agli altri un ingiusto profitto con altrui danno (art. 640).

L'intelligenza artificiale trova applicazione nello sviluppo dei sistemi antifrode in cui ormai da diversi anni vengono adottate tecnologie basate su reti neurali (accanto alle tipiche tecniche rule-based) in quanto il rilevamento delle frodi si riconduce tipicamente all'identificazione di comportamenti anomali, a similitudini comportamentali, correlazione di eventi, ecc. Questi sistemi permettono di analizzare grandi quantità di dati di diverse tipologie e, se il caso, generare un segnale di allerta denominato *bandiera rossa* (red flag).

Un sistema esperto può creare regole di ragionamento come queste:

1. SE (personaA AND personaB si conoscono) AND (incidente con auto di personaA AND auto di personaB) ALLORA bandiera=0,6;
2. SE (testimoneA AND testimoneB si conoscono) OR (testimoneA coinvolto in altri incidenti AND testimoneB coinvolto in altri incidenti) AND bandiera=0.5 ALLORA bandiera=0,8.

Queste semplici regole hanno la forma dei noti IF-THEN-ELSE; è possibile creare regole più complesse usando il ragionamento logico per tentare di riprodurre il modo di ragionare di un perito assicurativo.

L'approccio statistico basato sull'analisi delle anomalie rispetto ai normali comportamenti (*outlier detection*) è un efficace strumento per individuare qualcosa che non va. Per esempio, una neural network può prendere in input i dati di un movimento bancario e classificarlo come eventuale frode se è diverso dalle solite operazioni. oppure creare un modello del comportamento del cassiere, da cui generare un allarme se effettua pagamenti diversi dai soliti nel ciclo passivo.

Le neural network con addestramento supervisionato sono molto utili nella verifica della contraffazione del logo su un prodotto, grazie alla capacità di trovare il logo nell'immagine, anche se è ruotato o tagliato in parte, e poterlo classificare falso rispetto a una raccolta di logo corretti e falsificati.

Purtroppo, le neural network hanno il difetto di non rendere facilmente spiegabile come hanno ricavato un output a partire dall'input (il cosiddetto *problema della scatola nera*, già illustrato nel capitolo 31). Per cui, quando è rilevata una frode, difficilmente si può capire quale input in particolare ha attivato l'allarme.

La creazione di sistemi anti frode presenta un'ulteriore difficoltà molto importante: spesso ci sono pochissimi casi di esempio (mentre le neural network tipicamente richiedono tantissimi dati) e non si riesce neanche a delineare cosa cercare. Inoltre, chi subisce la frode non intende parlarne volentieri, rendendo difficile svolgere analisi, definire cosa cercare e creare scenari di test.

La valutazione delle prestazioni di un sistema anti frode deve porre attenzione ad alcuni indici specifici. Per esempio, nel caso del controllo delle transazioni bancarie, gli indici importanti sono:

- falso positivo: transizione segnalata come frode (positivo) ma in realtà lecita (falso), crea una perdita dovuta alla transizione non effettuata e fastidio o anche danno ingente al cliente;
- falso negativo: transizione non segnalata come frode (negativo) ma in realtà frode (falso), crea una perdita economica al cliente.

Può esserci qualche difficoltà nel trovare la letteratura tecnica aggiornata perché chi crea questi sistemi non vuole farne conoscere le caratteristiche ai possibili frodatori per evitare di fargli scoprire i punti deboli da sfruttare ed eludere i controlli.

44 L'IA per la sicurezza OT

Sebbene esistano molte tecnologie e soluzioni di IA applicate alla sicurezza per le reti IT (tecnologia dell'informazione), queste sono estremamente ridotte nell'applicazione alla sicurezza alle reti OT (tecnologia operativa solitamente adottata nel settore industriale).

Le reti OT sono state tradizionalmente isolate dalle reti IT e, pertanto, l'attenzione verso i parametri di riservatezza ed integrità è stata minore; inoltre, per natura, le reti OT sono più statiche, portando rischi legati all'obsolescenza (per esempio l'utilizzo di vecchi protocolli vulnerabili) dell'infrastruttura. Data però la tendenza crescente verso l'interconnessione delle reti e la divulgazione di attacchi sempre più sofisticati contro ambienti OT, gli operatori riconoscono sempre più la necessità di attuare soluzioni di sicurezza ampiamente note in IT anche in ambito OT.

In ambiente OT si evitano modifiche alla configurazione per ridurre il rischio di errori di configurazione o di effetti collaterali imprevedibili e dannosi per il funzionamento della rete. Pertanto, fino a ora, le tecnologie di sicurezza non intrusive sono risultate le favorite e l'analisi dei dati di rete raccolti in modo passivo è la tecnologia ritenuta più accettabile per le reti OT.

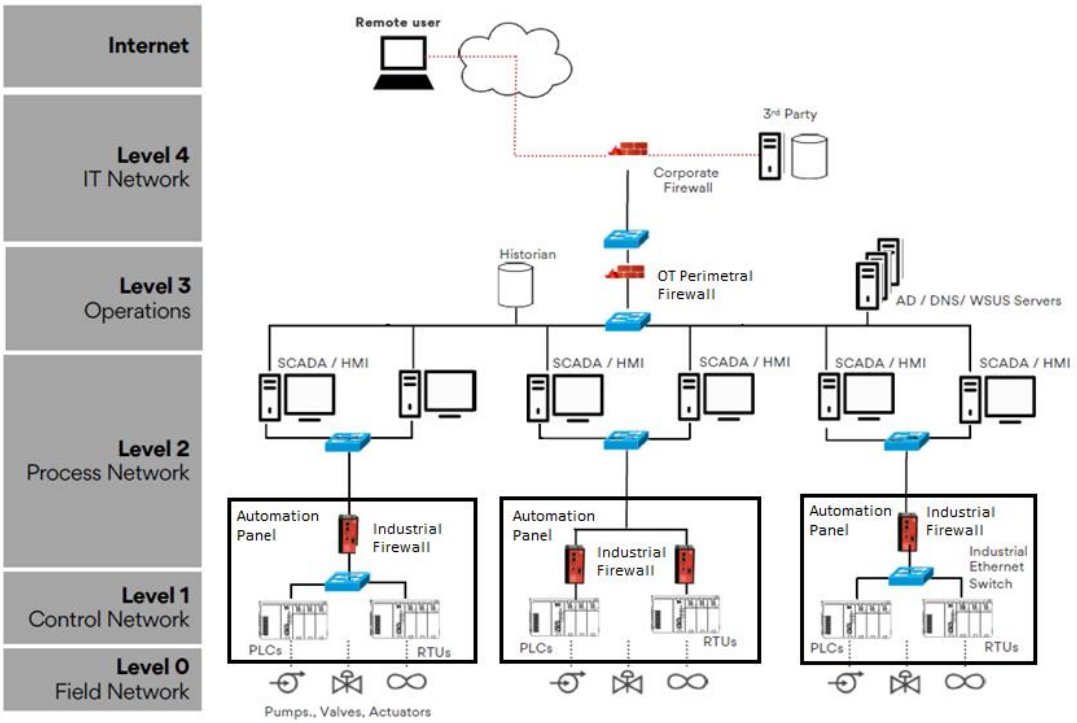


Figura 39 - Esempio di architettura OT con micro-segmentazione dei firewall

La cosa si complica se si considerano e si confrontano i paradigmi che regolano la sicurezza IT rispetto a quella OT.

Se in ambito IT la sicurezza si basa sul rispetto dei parametri RID (riservatezza, integrità, disponibilità), in ambiente OT l'ordine di questi tre fattori va letto al contrario e sempre correlato con gli aspetti legati alla sicurezza fisica: le caratteristiche irrinunciabili sono infatti la disponibilità (per le ovvie implicazioni legate alla continuità del processo o del servizio e la sicurezza fisica delle persone) e l'integrità (un comando di lettura scambiato con uno di scrittura in un PLC può causare danni enormi), mentre la riservatezza è quasi un parametro accessorio.

Altro elemento distintivo riguarda le competenze e la sensibilità del personale di impianto, di produzione e di manutenzione, in quanto non sempre adeguate per fronteggiare le potenziali minacce OT o semplicemente i frequenti (e costosi) incidenti che periodicamente si verificano. Analogamente, il personale IT, nella stragrande maggioranza dei casi, non ha la preparazione adeguata per gestire problemi legati all'automazione del processo produttivo (p.e. effettuare operazioni nei quadri elettrici presenti sulle linee di produzione).

Per questi motivi si stanno diffondendo nuove tecnologie, con caratteristiche adattative, in grado di individuare in tempo reale vulnerabilità nei dispositivi e nelle architetture, incidenti, attacchi, e potenziali minacce. Tecnologie in grado di segnalare le anomalie OT e, in alcuni casi, dotate di un certo grado di autonomia decisionale.

Attraverso algoritmi di machine learning è possibile identificare il comportamento dei vari dispositivi (PLC, DCS, SCADA, Historian, ecc.) e le interazioni che avvengono tra di essi durante il processo produttivo, permettendo la definizione di una "baseline" (la configurazione dei nodi e delle comunicazioni nella rete di impianto).

Una volta definita la baseline, è possibile identificare le anomalie che si verificano nelle comunicazioni tra i nodi ed eventuali anomalie nel processo (ad esempio aumento dei volumi di scambio dati o della frequenza degli accessi a un nodo).

Si è rilevata quindi la necessità di adeguare i sistemi di intrusion detection a un uso non solo sulle reti IT ma anche sui sistemi SCADA[222].

Recentemente sono state introdotte tecnologie (firewall industriali e interfacce intelligenti) che permettono di implementare strategie attive ma non invasive di protezione, ad esempio la microsegmentazione. Essa prevede la suddivisione o segmentazione di una rete in sottoreti relative a una macchina, una linea o una parte di impianto gestite da uno o più apparati di automazione che concorrono al funzionamento di quella specifica area produttiva. La segmentazione si può realizzare attraverso apparati (p.e. firewall) industriali che si auto-configurano, attraverso meccanismi di ML.

45 L'IA per la sicurezza fisica

45.1 Controllo accessi

L'IA permette di controllare gli accessi se integrata in meccanismi di riconoscimento biometrico, come specificato al paragrafo 42.4.

45.2 Monitoraggio degli ambienti

Un aspetto importante della sicurezza fisica è rappresentato dal monitoraggio, ad esempio, in aree metropolitane, centri commerciali, stazioni ferroviarie e aeroportuali. Esso può utilizzare tecnologie innovative che prevedono l'affiancamento di sistemi di videosorveglianza con software di riconoscimento facciale assieme a strumenti per l'analisi dei comportamenti.

L'uso in particolare di tecniche di riconoscimento facciale ha avuto di recente uno dei più massicci impieghi nel mondo del commercio al dettaglio negli Stati Uniti nella catena Rite Aid Corp[223], che ha installato questo tipo di telecamere intelligenti in ben 200 negozi.

[222] Yasakethu, Lasith & Jiang, J. Intrusion Detection via Machine Learning for SCADA System Protection. 2013. 10.14236/ewic/ICSCSR2013.12. doi: 10.14236/ewic/ICSCSR2013.12.
[223] https://www.reuters.com/investigates/special-report/usa-riteaid-software.

Con l'analisi dei comportamenti è possibile non solo analizzare comportamenti potenzialmente pericolosi, o comunque illeciti, in tempo reale attraverso telecamere, ma anche effettuare analisi di tipo predittivo in modo da segnalare eventuali pericoli potenziali prima che avvengano.

Società come la ShotSpotte[224] propongono soluzioni per avvisare le autorità in tempo reale in caso di sparatoria, indicando il punto in cui è avvenuta grazie al rilevamento dei suoni captati con algoritmi di machine learning. Nella stessa direzione vanno anche le ricerche condotte dai Cadre Research Labs per conto del National Criminal Justice Reference Service statunitense[225]. Altre società, come la Predpol, si spingono verso la prevenzione: in questo caso la soluzione analizza le serie storiche dei crimini commessi in una determinata zona e propone modelli per localizzare sul territorio aree con un'alta probabilità che si verifichi un evento criminoso. I ricercatori del Research Triangle Institute hanno sviluppato, invece, modelli per prevedere il rischio di recidiva per autori di reato in fuga se un mandato non viene eseguito[226].

46 L'IA contro le fake news

Così come l'IA è al centro dell'attenzione per la generazione di "fake news", si sta cercando di utilizzarla anche per il loro riconoscimento.

Tuttavia questa applicazione dell'IA comporta problemi, allo studio di diversi gruppi di ricerca, per esempio quello del Computer science and artificial intelligence laboratory (CSAIL) del MIT in collaborazione con il Qatar computing research institute (QCRI)[227].

Questi gruppi di ricerca stanno utilizzando vari metodi per analizzare i siti dei media, gli account Twitter associati, la reputazione della fonte, il traffico web e altri fattori, al fine di creare classifiche di alta, media e bassa veridicità.

Elementi considerati sono: il sito web, la struttura dell'URL, il testo stesso (ad esempio si applica l'IA per individuare i "machine-generated text").

Sono stati rilevati alcuni problemi. Per esempio, il fatto che un testo sia "machine generated" non implica che esso sia illegittimo, oppure che un'affermazione può essere corretta in un certo momento storico e non veritiera in un altro.

Il Gruppo del MIT spera che, in futuro, la combinazione del controllo dei fatti con i sistemi di difesa individuati renderà i modelli via via più robusti contro gli attacchi (o le "ondate di fake

[224] https://analyticsjobs.in/education/artificial-intelligence-crime-used-5-ai-detect-crime/.
[225] https://www.ncjrs.gov/pdffiles1/nij/252038.pdf.
[226] https://www.ncjrs.gov/pdffiles1/nij/252038.pdf.
[227] https://www.agendadigitale.eu/cultura-digitale/intelligenza-artificiale-contro-le-fake-news-lo-stato-della-ricerca/.

news"). Ciò sarà possibile attraverso il continuo miglioramento dei modelli esistenti, ottenuto sviluppando nuovi algoritmi e costruendo insiemi di dati che coprano più tipi di disinformazione[228].

47 L'IA e la sicurezza nazionale

Big data, IA e capacità di calcolo sono in forte crescita e costituiscono fattori rilevanti per la sicurezza nazionale dal momento che aprono nuove frontiere e nuovi scenari, tutti da esplorare, nella conduzione delle attività d'intelligence.

Questi fattori portano a:

1) ottimizzazione nella raccolta di grandi moli di dati, fortemente eterogenei, strutturati e non, a partire da quelli prodotti sui social network grazie all'uso di algoritmi IA dedicati;
2) automazione delle tecniche di analisi sfruttando ulteriormente tecniche di IA, p.e. per l'analisi semantica del testo, l'interpretazione e la catalogazione di immagini, il *face and object recognition*, l'individuazione di contenuti orientati alla *digital cognitive intoxication*, l'individuazione di contenuti di propaganda e radicalizzazione di matrice jihadista, ecc.

Secondo uno studio di una autorevole università americana, una considerevole parte di utenti del noto social network Twitter sono dei social bot il cui fine ultimo è quello di diffondere false notizie, con effetti devastanti, come nel caso in cui l'azione è tesa a influenzare e alterare l'esito di elezioni politiche. In generale i social bot possono essere utilizzati anche per disseminare il network con contenuti malevoli (malware). Proprio l'IA è una delle tecnologie che può venire in aiuto e limitare i fenomeni appena citati, ad esempio attraverso lo studio e realizzazione di algoritmi noti come supervised machine learning bot detection (SMLBD), comunque sempre mediati dall'azione umana per evitare di incorrere nella chiusura di account reali, ma non riconosciuti come legittimi.

Bisogna infatti prestare attenzione al fatto che l'applicazione delle suddette tecnologie trova riscontro in contesti estremamente delicati. Ci sono quindi aspetti che devono essere tenuti in debita considerazione, a partire dall'affidabilità degli algoritmi utilizzati. Ad esempio, nel tentativo di limitare il fenomeno della disinformazione digitale, è necessario chiedersi quanto gli algoritmi selezionati siano veramente capaci di riconoscere contenuti orientati alla propaganda e alla radicalizzazione senza limitare il diritto all'espressione di un pensiero o di un'idea.

[228] https://news.mit.edu/2019/better-fact-checking-fake-news-1017.

47.1 Autonomous cyber defense

L'autonomous cyber defense è un nuovo campo di ricerca nell'ambito della difesa del ciberspazio. Esso mira a realizzare sistemi di difesa intelligenti autonomi per proteggere le infrastrutture militari rispetto a minacce informatiche emergenti basate su meccanismi cognitivi automatici non gestiti da personale umano, come per esempio sistemi autonomi di attacco informatico (*autonomous weapon systems*) e malware intelligenti (*autonomous intelligent malware*).

L'evoluzione dei sistemi informativi a supporto delle missioni, con l'introduzione della *global information grid* che ha portato le piattaforme (come il sistema di information management, il sistema C4ISR, il sistema di comunicazione, i sistemi d'arma ed i sistemi di controllo IoT) a interconnettersi attraverso reti su larga scala, rende sempre più complessa la supervisione della sicurezza da parte di operatori umani.

Un rapporto del Dipartimento della difesa degli Stati Uniti[229] spiega infatti che l'identificazione delle vulnerabilità introdotte in sistemi complessi è estremamente difficile e "in un mondo perfetto, i sistemi operativi dei sistemi di difesa sarebbero in grado di dire a un comandante quando e se sono stati compromessi, se il sistema è ancora utilizzabile in modalità completa o degradata, identificare alternative per aiutare il comandante a completare la missione e infine fornire la capacità di ripristinare il sistema a uno stato noto e affidabile".

Nelle future missioni militari, i sistemi C4ISR (command, control, communications, computers, intelligence, surveillance and reconnaissance) utilizzati in circostanze di combattimento, i sistemi di comunicazione e i sistemi d'arma richiederanno capacità di difesa superiore rispetto a quella offerta dalle normali difese cyber come dispositivi firewall, IDS, e centri di controllo SOC. Avranno infatti necessità di disporre di capacità di difesa informatica intelligenti, autonome, realizzate attraverso agenti informatici artificiali, specializzati nella difesa informatica attiva (*autonomous intelligent cyber-defense agents*).

In questo contesto si collocano anche i sistemi di *cyber-supported situational awareness* che diventeranno parte dei sistemi informativi di comando e controllo delle operazioni militari. Queste piattaforme coordinando anche le attività degli agenti autonomi, introducono un meccanismo di supporto decisionale per le operazioni informatiche aiutando i centri di comando militare a comprendere le implicazioni del cyberspazio e proponendo piani di riparazione per ottenere la garanzia della missione.

Un agente autonomo è un sistema situato all'interno di un ambiente che ha cognizione dell'ambiente e agisce su di esso nel tempo, perseguendo la propria strategia per ottenere effetti nell'ambiente. Un agente è tipicamente associato a un certo contesto, interagisce con l'ambiente tramite sensori che forniscono input e attuatori che gli consentono di agire e influenzare quell'ambiente specifico agendo verso un obiettivo. Nel libro "*Artificial Intelligence - A modern approach*" Stuart Russell e Peter Norvig introducono una classificazione degli

[229] Office of the Under Secretary of Defense for Acquisition, Technology and Logistics. *Resilient Military Systems and the Advanced Cyber Threat*. USA: Defense Science Board, January 2013.

agenti che spazia da semplici agenti riflessi, dove non esiste un modello interno dell'ambiente, a agenti basati su modelli e obiettivi e agenti basati su funzioni di utilità.

L'architettura di riferimento[230] di un agente autonomo intelligente per la *cyber-defense,* sviluppata tra il 2016 e il 2019, prevede che il sistema autonomo sia in grado sempre di svolgere le seguenti azioni fondamentali:

- monitorare un perimetro di un sistema da difendere;
- rilevare attacchi informatici;
- elaborare piani di contromisure;
- eseguire tatticamente tali piani;
- riferire le loro azioni agli operatori umani;
- avere capacità di *autonomous learning* per essere in grado di contrastare adeguatamente minacce come malware le cui tattiche, tecniche e procedure evolvono nel tempo.

L'architettura ipotizzata dallo studio prevede, oltre ai moduli di raccolta informativa, normalizzazione ed elaborazione, anche un modulo destinato alla memorizzazione della conoscenza (knowledge base), che contiene le informazioni inerenti il modello dell'ambiente, la descrizione dello stato, delle strategie delle minacce informatiche note e delle attività passate. Il modulo si configura dunque come un database a grafo le cui relazioni sono definite da un'ontologia (spesso compatibile con lo standard STIX 2.1)[231] espressa tramite un linguaggio strutturato per la modellazione del dominio, realizzando un grafo di conoscenza alimentato dai moduli di acquisizione e dalle sorgenti informative.

L'agente cognitivo, dunque, attraverso appositi sensori raccoglie le informazioni dal campo e dai sistemi componenti la missione, le correla con le informazioni della knowledge base al fine di identificare il rischio associato e, in caso di minaccia rilevata, avvia i processi di pianificazione, selezione ed esecuzione della risposta, orchestrando le attività di risposta con altri agenti cognitivi impiegati nel medesimo scenario, mediante le funzionalità di collaborazione.

Nell'espletamento delle funzionalità di identificazione della minaccia e di selezione ed esecuzione della risposta, l'agente cognitivo utilizza algoritmi di artificial intelligence quali anomaly detection, reasoning e reinforcement learning. Tali metodi però hanno il limite di ridurre il processo decisionale a un solo passaggio. Tipicamente gli attaccanti, in un quadro di attacco complesso, pianificano diverse mosse e successive reazioni alle mosse da parte dell'avversario. Al fine di consentire agli agenti di prendere decisioni efficaci nell'ambito della battaglia, occorre introdurre un nuovo modello decisionale[232]. Attualmente la ricerca sta verificando la convergenza di diverse correnti di lavoro come architetture cognitive (ad esempio l'adaptive control of thought-rational, ACT-R), il processo decisionale naturalistico, insieme con la teoria dei giochi e il machine learning.

---

[230] A. Kott, P. Theron, M. Drašar, E. Dushku, B. LeBlanc, P. Losiewicz, A. Guarino, L. V. Mancini, A. Panico, M. Pihelgas and K. Rzadca. *Autonomous Intelligent Cyber-defense Agent (AICA) Reference Architecture, Release 2.0.* USA: US Army Research Laboratory, 2019.

[231] https://docs.oasis-open.org/cti/stix/v2.1/cs01/stix-v2.1-cs01.pdf

[232] Dr. Paul Théron, Dr. Alexander Kott. When Autonomous Intelligent Goodware will Fight Autonomous Intelligent Malware: A Possible Future of Cyber Defense. *Proceedings of the Military Communications Conference, MILCOM-2019, Norfolk, VA: November 12-14*, 2019.

Tra le strategie di difesa applicabili dall'agente, vi è la "*difesa informatica attiva*" (active cyber defense), ossia un'azione difensiva intrapresa per distruggere, annullare o ridurre l'efficacia delle minacce informatiche contro forze e risorse amiche. In questo scenario gli agenti cognitivi sono configurati per organizzare dinamicamente strategie di attacco sulla base delle informazioni del target e delle tattiche, tecniche e procedure contenute all'interno di una knowledge base. La selezione di tali strategie avviene sempre attraverso processi cognitivi che considerano strategie a più fasi e imparano dall'esperienza pregressa.

Diversi programmi di sviluppo sono stati avviati per la realizzazione di tali agenti. Tra le iniziative si cita il progetto Cyber Trainer[233] (finanziato dalla Regione Abruzzo) che ha come obiettivo la realizzazione di una piattaforma di *cyber range*, ossia una piattaforma per sviluppare esercitazioni di tipo *cyber wargame*. In tale piattaforma, un agente cognitivo intelligente è in grado di simulare una squadra di attacco (*red team*), ossia di effettuare attacchi cyber basandosi sulla propria intelligenza e conoscenza pregressa, operando in maniera automatizzata attacchi complessi e variati che tengono in conto le azioni e le repliche della squadra di difesa (blue team)

L'agente è composto da un modulo di selezione delle strategie di attacco (*AI strategy module*) sulla base di tecniche di reinforcement learning. Vi è poi il modulo di *attack path selector* che, in base allo stato corrente dell'ambiente, espresso in termini di configurazione delle macchine target e raggiungimento di obiettivi intermedi, indica quanto ciascuna azione intrapresa avvicini l'agente al raggiungimento del suo obiettivo. L'agente fornisce per ciascuno stato una lista di azioni eseguibili al modulo di orchestrazione che, a sua, volta esegue la tattica scelta. La tattica potrà essere rivista dall'agente in base alle informazioni che dispone dell'ambiente ed alle azioni della squadra di difesa.

L'agente sviluppato nella versione prototipale è in grado, dunque, di:

- emulare in modo automatico il comportamento di un membro di una squadra di attacco (o di un'intera squadra);
- analizzare e valutare l'ambiente della rete bersaglio dell'attacco;
- scegliere la migliore strategia per raggiungere uno o più obiettivi prefissati;
- attuare più tattiche in parallelo per massimizzare il risultato.

Nella fase di addestramento l'agente apprende la migliore strategia di transizione mediante una funzione in grado di selezionare la "prossima migliore azione" basandosi sullo stato corrente e sul rinforzo, al fine di raggiungere l'obiettivo prefissato. Gli algoritmi di reinforcement learning utilizzati sono attualmente: Q-Learning, SARSA,DQN (Deep Q-Learning Network), Actor-Critic, A3C.

Il piano di sviluppo prevede la realizzazione di funzionalità di integrazione degli agenti con l'intelligence per ottenere informazioni di contesto più adeguate al contrasto di nuovi malware, capacità di coordinamento con agenti esterni alle missioni riconosciuti come amici, capacità di difesa in particolare verso minacce di tipo malware.

[233] https://www.leonardocompany.com/it/press-release-detail/-/detail/cyber-trainer-abruzzo

- SETTIMA PARTE: CONCLUSIONI

48 Uno sguardo al futuro dell'IA per la sicurezza

Chi lavora nella cybersecurity è sempre più attento a considerare la possibilità che innovazioni *disruptive*, come quelle portate dalle tecniche di IA e in particolare di machine learning, possano risolvere annosi problemi come la mancanza di staff, i budget insufficienti, la continua crescita della gravità degli attacchi e la crescente complessità di gestione di numerosi ambienti.

Oggi sono numerosi gli ambiti della ricerca su soluzioni IA per la cybersecurity che potrebbero avere in futuro un grande impatto sulle modalità con cui viene concepita e progettata la difesa degli ambienti ICT. In sintesi, l'IA sta dimostrando di avere il potenziale per:

- incrementare la complessiva "affidabilità" dei sistemi ICT;
- identificare (anche in modo predittivo) rischi e minacce;
- progettare una difesa e una risposta molto più rapide ed efficaci (ad esempio, attraverso correlazioni intelligenti e risposte adattative);
- semplificare e ridurre il lavoro per gli addetti alla cybersecurity;
- migliorare l'interazione degli utenti con la tecnologia;
- aumentare la consapevolezza delle persone, la visibilità complessiva sulla situazione, anche di tipo prospettica.

Come riportato nel rapporto "Artificial intelligence and cybersecurity: opportunities and challenges" (pubblicato nel marzo 2020 dal NSTC[234]), sintesi di un incontro sulle opportunità nascenti nell'intersezione degli ambiti cybersecurity e intelligenza artificiale, obiettivo della ricerca sarà nei prossimi anni quello di verificare l'applicabilità in contesti pratici di quanto viene sviluppato, tenendo presenti i limiti che l'utilizzo dell'IA potrebbe dover subire.

Ad esempio, per quanto riguarda l'incremento dell'affidabilità complessiva dell'IT, un ambito promettente è quello che punta all'utilizzo di sistemi IA per realizzare programmi più sicuri (vedere paragrafo 42.9). Le tecniche IA sono usate per individuare errori nei programmi, per indicare best practice, per identificare vulnerabilità. I cosiddetti "coding partner" sono algoritmi IA che assistono gli sviluppatori nel comprendere ambienti software complessi prima di metterci mano, che verificano la robustezza delle modifiche prima che siano attuate. Inoltre, una volta che il software è creato, può essere compito di un agente IA individuare eventuali vettori d'attacco, ricercare errori di configurazione che possono portare a difetti, monitorare reti o suggerire miglioramenti per la sicurezza del software.

Con riferimento all'identificazione predittiva di rischi e minacce, si è già parlato in precedenza (capitolo 42) dell'uso oramai maturo dell'IA per accrescere le capacità dei sistemi antimalware, di sicurezza della rete e quant'altro. Un ambito, però, che vedrà importanti sviluppi in futuro è quello della difesa da attacchi che faranno a loro volta uso di tecniche IA (vedere il capitolo

[234] https://www.nitrd.gov/news/2020/artificial-intelligence-cybersecurity.aspx.

39). Gli attaccanti saranno presto in grado di utilizzare sistemi intelligenti per ottenere informazioni sui sistemi oggetto di attacco, per capire come eseguire un attacco efficace, quali sono i punti deboli da colpire, come agire senza essere riconosciuti. Servirà quindi una difesa in grado di resistere a questi "attacchi e analisi automatiche" effettuate da esterni. Tecniche che si stanno rivelando promettenti sono quelle dell'"isolamento automatico", della "defensive agility" (ossia, utilizzo di simulazioni e aggiornamenti per rafforzare le difese), delle strategie "mission-specific" che utilizzano la competenza di esperti di dominio per categorizzare attacchi e risposte.

Un ambito che vedrà grande sviluppo in futuro è poi quello della "autonomous cyber defense" (vedere paragrafo 47.1), che dovrà essere pensata però non più per singole soluzioni, ma per un disegno complessivo della difesa. Come noto, l'IA sta già oggi dimostrando che le prestazioni di singoli strumenti di cybersecurity possono essere potenziate notevolmente: in futuro servirà un migliore coordinamento e un'orchestrazione dei diversi strumenti, un modello che verifichi le rispettive interazioni, che includa la supervisione umana e che punti a obiettivi più ampi. Ad esempio, il riconoscimento di un attacco in corso fin dalle sue prime fasi o la predizione di come evolverà l'attacco sulla base dell'analisi dei segnali di early-warning e quindi l'attivazione delle misure preventive. O, ancora, l'adattamento della difesa sulla base dei cambiamenti nei contesti esterni.

Il consiglio degli esperti (considerando questi sviluppi dell'IA per la cybersecurity) è che si devono comprendere sempre bene tutte le possibili conseguenze di un falso positivo e falso negativo. Molta attività di ricerca sarà necessaria nei prossimi anni per analizzare nel dettaglio – identificando i rischi e utilizzando tecniche di validazione incrociata – perché l'uso dell'IA per la cybersecurity non possa a sua volta essere fonte di ulteriori problemi.

Pensando al futuro, un ruolo importante potrebbero giocarlo le startup e la tecnologia che sempre più si sta orientando verso il cloud e verso lo sviluppo di app integrabili, spesso basate su microservizi.

Per fare un esempio a questo proposito, citiamo la soluzione della startup Spitch, società nata in Svizzera ma presente in altri stati europei fra cui l'Italia, in grado, fra le altre cose, di effettuare l'autenticazione utilizzando il timbro della voce. L'utilizzo della voce come dato biometrico, rispetto ad altri più tradizionali (come ad esempio l'iride o il riconoscimento facciale), consente, nel caso di conversazioni prolungate, di aggiungere anche un ulteriore controllo sullo stato d'animo di chi parla, riuscendo ad esempio a prevenire eventuali illeciti (che possono essere traditi dal tono di voce).

Questa soluzione può essere integrata con soluzioni di SCA (strong customer authentication): l'autenticazione può includere il riconoscimento vocale.

Si tratta solamente di un esempio, anche se alcune aziende hanno già iniziato a ragionare su questo tipo di possibilità e sui potenziali benefici.

Analizzando studi pubblicati, ad esempio quello di Forbes "Top 10 Cybersecurity Companies To Watch In 2019"[235], si vede come molte delle startup che operano in ambito sicurezza in realtà siano focalizzate proprio su tematiche legate all'intelligenza artificiale, che quindi resta uno dei temi guardati con più attenzione e su cui sicuramente ci saranno evoluzioni interessanti nei prossimi anni.

Intervista a Raoul Brenna, Responsabile Security by Design e Cybersecurity Awareness, Fastweb

Si occupa di sicurezza informatica da più di 10 anni, affrontandola sia da un punto di vista "tradizionale" (assessment tecnologici e di processo, elaborazione di linee guida, definizione di policy, security governance), sia toccando ambiti maggiormente innovativi (studio dei nuovi trend di attacco e minaccia, identificazione di rischi e opportunità in relazione a nuove tecnologie, sicurezza del fattore umano, sicurezza IoT e definizione di soluzioni sostenibili anche per il "legacy") e promuovendo un approccio costantemente rivolto alla sperimentazione, come importante veicolo di consapevolizzazione a tutti i livelli. Oggi in Fastweb guida la funzione che si occupa di affrontare proattivamente e "by design" gli aspetti di cybersecurity nelle nuove iniziative ICT, oltre a curare la sensibilizzazione aziendale sul tema.

Domanda 1. Quali sono gli ambiti di applicazione dell'intelligenza artificiale nella cybersecurity?

In linea di principio, ipotizzo una suddivisione concettuale in tre macro-ambiti di applicazione e adozione:

- tutto ciò che riguarda l'identificazione (anche predittiva) di fenomeni;
- l'aiuto a effettuare correlazioni "intelligenti" (nei limiti di quanto le IA possano definirsi tali) e a gestire risposte adattative;
- la semplificazione dell'interazione tra tecnologie e persone, secondo molteplici prospettive.

Provo a dare alcuni elementi sintetici per ciascuno di questi tre filoni, identificando quelli che possono essere le declinazioni più promettenti di ciascuno.

Vi è innanzitutto l'ambito dell'identificazione. Con questo intendo riferirmi a tutto ciò che riguarda il riconoscimento tempestivo di fenomeni che, pur riconducibili ad un pattern noto, se ne discostano in misura anche significativa. Anche gli aspetti legati alla contestualizzazione vengono esaltati dalla tecnologia, data la capacità di questi algoritmi di introdurre

[235] https://www.forbes.com/sites/louiscolumbus/2019/06/16/top-10-cybersecurity-companies-to-watch-in-2019/#3cc82f766022.

nell'elaborazione la possibilità di discriminare il contesto e di operare valutazioni di conseguenza.

Nella pratica, al di là di tutte le possibili sfumature che si possano tentare di introdurre, le applicazioni più concrete in questo primo specifico ambito paiono essere legate a:

- identificazione (o rilevamento in assenza di signature o euristiche note) di nuovi malware ed in generale nuove minacce;
- monitoraggio dell'attività standard di sistemi e persone per rilevare deviazioni e anomalie potenzialmente malevole (anche in assenza di una definizione di "anomalia"); l'esempio classico è l'identificazione di traffico generato da botnet;
- rilevamento del phishing; qui, le tecnologie possono operare discriminando la tipologia in senso ampio sull'intero messaggio o, in modo più mirato, analizzandone specifiche caratteristiche.

Un passo ulteriore (seconda macro-area) potrà essere quello di abilitare decisioni autonome nell'ambito dello scopo prestabilito. Con un livello di complessità incrementale, penso a:

- ricerca e analisi rapida di correlazioni tra eventi (log, ecc.), anche "sparsi" nell'infrastruttura ICT e/o di natura e formato differente;
- riconfigurazione automatica (self-patching) di reti e sistemi, in relazione alla scoperta di vulnerabilità;
- per i più "sperimentatori" tra i gestori di infrastrutture ICT e sistemi di cybersecurity, il futuro promette la disponibilità di honeypot dinamiche.

Questi punti sostengono la nozione oggi sempre più propagandata di "cognitive SOC" o altri approcci similari.

Infine, una terza macro-area di interesse per l'applicazione di IA/ML alla security, in realtà già nota e per certi aspetti sviluppata da almeno un decennio, è quella che prevede l'analisi dei dati e dei comportamenti legati a (o pensati per) le persone. Parliamo qui di comportamenti e dati non strutturati, messi in relazione con contesti di riferimento e informazioni note e generate da sistemi. Il tutto a supporto delle attività di chi deve "lavorare con i dati". Nella pratica si possono sviluppare le più disparate applicazioni, tra cui citiamo a titolo di esempio:

- l'interpretazione di dati non strutturati (creati "dalle persone per le persone") e messa in relazione con dati strutturati (prodotti da sistemi) per scoprire nuovi fenomeni;
- l'autenticazione adattativa, basata su scenari e indicatori di rischio dinamici e contestualizzati; il traguardo (che svariati attori primari indicano come raggiungibile nei prossimi anni) è quello di un mondo "passwordless";
- il rilevamento delle frodi, di fatto caratterizzato da operazioni concettuali come: identificazione pattern anomali, riconduzione a pattern noti, similitudini, relazioni, ecc.

Domanda 2. I malintenzionati, invece, quali usi potrebbero farne per i propri attacchi?

Le tecniche di elaborazione dei dati basate su IA/ML, se adottate in ottica offensiva, mi pare aprano la strada a due macro-scenari principali nell'ambito stretto della cybersecurity, ai quali se ne affianca uno più generale e più preoccupante:

- applicazione al phishing e alla manipolazione dei comportamenti;
- utilizzo ai fini di potenziamento del malware;
- in un contesto che si delinea sempre più come "cyber-fisico", manipolazione della "realtà".

Ancora una volta, provo a caratterizzarli meglio.

Per quanto riguarda il primo tema, nonostante l'efficacia delle campagne generaliste rimanga elevata, gli attaccanti sono sempre alla ricerca del "phishing perfetto". Quello che in gergo si chiama "spear-phishing", ossia phishing estremamente contestualizzato, e che richiede un notevole dispendio di tempo ed energie per la preparazione. Un supporto automatizzabile, scalabile e che produca risultati "ragionevoli" (e migliorabili nel tempo mediante autoapprendimento) certamente può accrescere in modo preoccupante l'efficacia e la portata di questo tipo di attacchi. I principali ambiti di intervento e di supporto per l'IA/ML che si vanno quindi delineando in quest'area sono sostanzialmente:

- riconoscimento dei target più "promettenti";
- predisposizione al pagamento del riscatto sulla base di dati storici/statistici;
- generazione di URL credibili nelle email di phishing (anche con "typosquatting");
- creazione di tweet/post attrattivi e potenzialmente sincronizzati con l'attività del target (destinatario o impersonato) sui social.

Nel futuro si intravede una generalizzazione di questo approccio, automatizzando la creazione di siti e indirizzi email credibili nella loro interezza.

Anche rispetto al miglioramento dei sistemi di attacco, la IA può giocare un ruolo importante: l'introduzione di modifiche opportune ai codici, atti a renderli meno riconoscibili dalle soluzioni antimalware, e di comportamenti variabili nel malware in relazione alle difese rilevate sul target aumentano l'efficacia degli attacchi. Più in generale parliamo di:

- kit di attacco con comportamenti "guidati da IA";
- offuscamento dei comportamenti e del codice finalizzato all'evasione delle tecniche di signature-based detection e di rilevamento euristico;
- implementazione di "next generation C&C" per le botnet, così che i bot possano avere le sole "armi" che occorrono quando occorrono.

Quelli esplorati sopra sono tuttavia scenari e casi di applicazione molto settoriali e legati agli ambiti tradizionali della cybersecurity. Occorre tuttavia considerare che, alla luce della crescente pervasività dell'ICT nella realtà di tutti i giorni (negli oggetti, nelle interazioni, ecc.), il potenziale di manipolazione della realtà stessa diventa enorme. Parliamo di infrastrutture

critiche, di settore sanitario, di accesso a servizi pubblici e di tutto ciò che sempre più caratterizza la vita del cittadino.

Su molti di questi ambiti, dove la tecnologia è entrata in modo quasi prepotente prima che si sviluppasse una matura attenzione ai temi di cybersecurity, l'aspetto preoccupante è che al giorno d'oggi elaborazioni basate su IA/ML possono essere svolte in tempo reale anche da oggetti di potenza relativamente limitata che è possibile nascondere nelle installazioni fisiche. Questo permette di ampliare in misura notevole le già ampie possibilità di sfruttamento delle numerose falle (di sicurezza logica ma non solo) che spesso caratterizzano questo tipo di facility.

Domanda 3. Quali limiti stanno emergendo con i primi utilizzi dell'IA nella cybersecurity?

Non mi soffermo eccessivamente sulla premessa, che penso sia oramai nota ai più, che l'intelligenza artificiale non è "realmente intelligente". Ci sono molti esempi di pubblico dominio in cui si mostra come con pochi elementi di trucco facciale, o di nastro adesivo colorato su cartelli stradali, si alterano le funzionalità dei sistemi di face recognition o di guida autonoma.

Per restare più nel mondo digitale, analoghi esempi mostrano come, con l'applicazione di pattern di rumore praticamente indistinguibili all'occhio umano, si possano alterare completamente le prestazioni degli algoritmi di classificazione delle immagini. Le IA "ragionano" secondo paradigmi differenti (e spesso black box, fortemente legati all'addestramento ma poco o per nulla comprensibili nelle logiche intrinseche). Quindi quando falliscono, spesso si tratta di "epic fail". Questo perché le IA emulano il comportamento umano nelle condizioni "a regime", ma le loro scelte nelle situazioni meno coperte dal training o non accuratamente modellate possono divergere significativamente dalle attese… e dalla "ragionevolezza".

Ciò apre la strada a tecniche di attacco per il sovvertimento delle IA, con introduzione di "bias" nei dati di training, alterazione della percezione di "normale" mediante manipolazione dei dataset (the danger of small changes), e in generale alterazione dei comportamenti. Sono allo studio per questi attacchi dei remediation a livello teorico, ad esempio attraverso l'analisi dell'entropia degli stati perturbati (l'assunto è che in qualche modo sia possibile discriminare le perturbazioni derivanti dal "rumore" del dato reale, da quelle volutamente introdotte da un attaccante) o altri approcci comunque estremamente complessi, la cui efficacia deve essere compresa.

L’adozione dell’IA (con particolare riferimento al contesto enterprise e delle infrastrutture critiche) dovrà a mio avviso considerare alcuni insegnamenti che le esperienze attuali fanno emergere.

- L'output dei sistemi di IA (ed in particolare di quelli basati su machine learning) è rappresentato da una probabilità statistica. Non esistono gli assoluti. Sono assolutamente da mettere in conto falsi positivi, complessità del tuning delle soluzioni (perché siano realmente efficaci nel contesto), difficoltà nel triage (a fronte di un'attesa maggiore velocità e migliore approssimazione nella detection iniziale). Non si può prescindere dalla messa in campo di team "esperti" (di dominio applicativo e in ambito della "data science") per la massimizzazione del valore.
- Gli approcci guidati da IA/ML si adattano bene a nuove varianti di minacce conosciute, meno bene alla comparsa di vettori di attacco completamente nuovi, pertanto il supporto dell'analista rimane insostituibile in questo secondo caso.

Se ci si chiedesse se si intravede in modo consistente un movimento nella direzione di sviluppo di "algoritmi autonomi intelligenti", una prima risposta potrebbe arrivare dal fatto che "*Facebook recently abandoned an AI experiment after ‘chatbots’ invented their own language which was not understandable by humans*".

Quest’ultima è chiaramente una piccola provocazione, ma non così lontana dalla realtà. Una lezione interessante può derivare da questo: data la necessità di allontanare le persone dalle aziende remotizzando il lavoro a causa dell’epidemia di coronavirus, Facebook ha affidato il compito di selezionare e rimuovere i contenuti offensivi o inappropriati, solitamente svolto da svariate centinaia di moderatori, ad algoritmi IA/ML che nell’operatività quotidiana si limitavano a supportare i moderatori[236]. Risultato? “*Multiple reports emerged that Facebook was falsely flagging and removing legitimate content from sites such as the BBC, Buzzfeed and USA Today, as well as users’ posts*”. La stessa Google, per citare un altro player fortemente coinvolto nell’uso di queste tecnologie, rileva questo quasi contemporaneamente (in merito ai contenuti su YouTube, su cui pure si è avuta una riduzione del presidio umano): “*With fewer people to review content, our automated systems will be stepping in to keep YouTube safe. More videos will be removed than normal during this time, including content that does not violate our Community Guidelines.*”

La discriminante maggiore penso sia che finché le operazioni guidate da IA/ML restano “aiuti” e agevolazioni agli operatori, e questi ultimi sono in grado di comprendere e governare i razionali di determinate scelte, il beneficio potrà essere sfruttato al massimo.

236 https://www.forbes.com/sites/kateoflahertyuk/2020/03/18/covid-19-fallout-why-is-facebook-wrongly-removing-legitimate-content

49 Raccomandazioni finali all'utilizzo dell'IA per la sicurezza informatica

In questa sezione conclusiva forniamo alcune raccomandazioni sull'utilizzo dell'IA per la sicurezza informatica.

Abbiamo suddiviso le raccomandazioni in tre gruppi:

1. raccomandazioni per le organizzazioni;
2. raccomandazioni per gli sviluppatori di sistemi di IA;
3. raccomandazioni per il legislatore.

49.1 Raccomandazioni per le organizzazioni

Iniziamo con alcune raccomandazioni dirette alle organizzazioni.

1. Creare un **gruppo di persone** operative con competenze multidisciplinari.
Occorrono diverse capacità professionali per gestire correttamente i vari aspetti che influenzano le scelte sulla sicurezza. Servono persone in grado di comprendere: il contesto in cui opera l'azienda, i vari livelli di difesa esistenti, il tipo di dati da trattare, gli ultimi aggiornamenti tecnologici.
Occorre anche individuare i ruoli manageriali con responsabilità precise relative all'organizzazione, governo e monitoraggio degli aspetti di sicurezza dei servizi intelligenti.

2. Effettuare periodici **riesami interni** delle misure di sicurezza.
Le misure vanno verificate e aggiornate con adeguata pianificazione temporale, per essere sempre al passo con gli attacchi possibili, anche tramite audit esterni.

3. Mantenere l'**uomo al centro** dei processi decisionali.
L'operatore umano non può essere completamente tenuto fuori dal processo decisionale, in un contesto così sofisticato come la difesa da attacchi informatici. Il sistema basato su IA deve sempre permettere di capire sta succedendo e l'intervento di operatori umani. Bisogna anche conoscere quali potenziali errori può compiere l'operatore umano e con quali impatti, in modo da pianificare l'opportuna formazione e rimediare ad eventuali errori.

4. Coltivare le **relazioni con la direzione**.
Occorre avere il supporto e l'appoggio della direzione quando sono necessari investimenti in sicurezza, facendo leva sulla fiducia acquisita dai recenti sistemi basati su IA e sulla corporate social responsibility come valore aziendale che sostiene la sicurezza dei servizi e applicazioni intelligenti.

5. Condurre l'**analisi di rischi**.
Un approccio sempre valido nell'ambito della gestione di sicurezza consiste nel valutare il proprio rischio e quanto si vuole ridurlo con strumenti basati su IA, per capire meglio quali risorse si possono dedicare nei limiti di competenze e budget disponibili.

6. Verificare il livello di **competenza degli operatori** di sicurezza informatica incaricati di gestire i sistemi basati su IA.
Il personale operativo deve essere in grado di organizzare in maniera adeguata le informazioni da analizzare e di interpretare correttamente i segnali di allarme ricevuti. Vanno considerati eventuali corsi di formazione per integrare le competenze necessarie.

7. **Non farsi ingannare** dalle belle parole e dalle grandi promesse.
Le funzioni aziendali dedicate agli acquisti dei sistemi di sicurezza devono aumentare la loro capacità di dialogare con le aziende venditrici per evitare di credere facilmente ad annunci pieni di entusiasmo e dal fascino fantascientifico e per non acquistare soluzioni costose solo perché possiedono elementi di IA.

8. Il **principio di proporzionalità** è sempre valido.
Solo l'importanza, la quantità e la particolarità dei dati che vengono trattati dall'organizzazione devono guidare l'adozione dell'IA. Per la difesa di certi contesti piuttosto semplici la IA potrebbe diventare un oggetto costoso, lento, inefficace e pesante da gestire.

9. Seguire il **principio di cautela nell'adozione** di una nuova soluzione.
Non bisogna introdurre un nuovo sistema solo perché è di moda o si sente dire in giro che ha prestazioni migliori o lo hanno adottato anche concorrenti e conoscenti. Un sistema nuovo può mostrare più vulnerabilità rispetto a un sistema già ben collaudato e aggiornato, specialmente quando viene adottata una tecnologia complessa come l'IA. Meglio eseguire una fase di affiancamento con sistemi già esistenti e una accorta fase di test.

10. Fare **attenzione ai dati** (se entra spazzatura può uscire solo spazzatura).
L'IA è un motore il cui carburante sono i dati, tanti e di ottima qualità, in cui vale ancora di più la classica regola GIGO: garbage in - garbage out, ovvero se entra spazzatura può solo uscire altra spazzatura. Occorre quindi impegnare molto tempo nell'adeguata preparazione dei dati per l'addestramento e la validazione dei sistemi basati su IA e avere indicazioni sulla tipologia di dati attesa durante la messa in esercizio.

11. Considerare l'importanza di **porsi le giuste domande**.
Quando si ragiona sull'introduzione di sistemi dotati di IA occorre rispondere a domande come le seguenti:

- Perché dovrei usare IA? Quali vantaggi mi offre?
- Ho dati sufficienti?
- Ho risorse adeguate?
- Come trovare chi mi aiuta a installare e configurare il sistema basato sull'IA?
- Quanto costa e quanto fa risparmiare?
- Quali aree coinvolgere?
- Qual è il freno all'adozione delle soluzioni?
- Chi è responsabile di un errore fatto dal software o dall'algoritmo?

- Chi è il proprietario del dataset utilizzato?
- Il sistema può adattarsi rapidamente a cambiamenti di contesto?
- Il sistema funziona bene anche se aumenta rapidamente la quantità di dati?

49.2 Raccomandazioni agli sviluppatori dei sistemi IA

Passiamo ora ad alcune raccomandazioni per gli sviluppatori dei sistemi di IA.

1. Diminuire il **problema del black box** per mostrare come ha ragionato il sistema.
I sistemi più noti di IA difficilmente spiegano perché forniscono una specifica risposta. Questo può invece essere importante quando occorre spiegare cosa è successo, per esempio nel ricostruire un certo attacco. Occorre approfondire le tecniche di trasparenza dell'IA per porre rimedio a questo difetto.

2. Considerare fasi di **test più specifici** per il contesto della sicurezza.
Esistono molteplici approcci per verificare le prestazioni di un sistema con IA, ma sono validi per un contesto generico. Il caso specifico della cyber security richiede una valutazione più accurata.

3. Attenzione nell'**evitare pregiudizi e distorsioni** anche involontari.
Sono noti i casi di cronaca in cui sistemi di IA presentano pregiudizi e distorsioni. Nell'ambito della sicurezza, essi possono essere sfruttati da un malintenzionato per superare le difese.

49.3 Raccomandazioni per il legislatore

Concludiamo infine con alcune raccomandazioni per il legislatore.

1. Creare una **certificazione del prodotto di sicurezza** con IA.
Per controllare i potenziali punti deboli sfruttabili da malintenzionati, per facilitare l'integrazione con soluzioni già installate e per evitare uno spreco di risorse è auspicabile la promozione di schemi di certificazione dei prodotti di sicurezza contenente IA, per esempio nel settore biomedicale, considerati le potenziali conseguenze sulle salute umana.

2. Creare un **percorso formativo specifico** per l'IA nell'ambito della sicurezza.
Negli ultimi anni sono nati decine di corsi di studio, pubblici e privati, per fornire competenze nell'ambito dell'IA. Alcuni di questi corsi sono nati solo per cavalcare l'onda del momento, piuttosto costosi, rivolti a studenti di variegata tipologia e con un corpo docenti non specializzati nella formazione. Un altro punto di attenzione riguarda il programma didattico, che non dovrebbe mirare solo alla immediata programmazione in un certo linguaggio, ma dovrebbe innanzitutto fornire strumenti di ragionamento e la creazione di solide basi su cui

costruire le frequenti evoluzioni. Si raccomanda pertanto un maggior controllo di tali corsi attraverso opportuni percorsi di qualifica.

3. Creare attenzione alle **implicazioni etiche** nell'opinione pubblica.
La Commissione Europea (UE) ha costituito un gruppo di lavoro per l'etica nell'uso dell'IA[237] che ha proposto le seguenti raccomandazioni, che dovrebbero essere ulteriormente promosse a livello nazionale[238]:

- tutelare i cittadini dall'uso improprio delle tecnologie intelligenti;
- assicurare che l'IA rispetti i diritti fondamentali delle persone e i principi delle nostre società;
- rispettare i principi base di liceità, correttezza, specificazione delle finalità, proporzionalità del trattamento, protezione dei dati fin dalla progettazione (privacy by design) e protezione per impostazione predefinita (privacy by default), responsabilità e dimostrazione della conformità (accountability), trasparenza, sicurezza dei dati e gestione dei rischi;
- evitare pregiudizi ed effetti discriminatori, come quelli basati sulla differenza di genere o sulle minoranze etniche;
- rendere l'IA tecnicamente solida e affidabile per evitare danni non intenzionali;
- valutare preventivamente i possibili rischi con un approccio di tipo precauzionale.

I principi essenziali forniti dalle linee guida sono sette:

- ci deve essere sempre un controllo umano;
- gli algoritmi devono essere affidabili, sicuri e resistenti di fronte alle incoerenze;
- i cittadini devono essere sempre informati sull'uso dei loro dati personali, oltre ad averne il pieno controllo;
- deve essere garantita la tracciabilità dei sistemi;
- dev'essere garantita la diversità e la non discriminazione;
- si deve lavorare a favore del benessere sociale e ambientale;
- deve essere attribuibile una responsabilità a chi gestisce questi particolari sistemi.

4. Attuare **azioni governative** per lo sviluppo dell'IA.
Sfruttare lo sviluppo della IA nell'ambito della sicurezza informatica come volano di crescita e trasformazione digitale per l'intero Paese.

Per ricordare l'importanza delle azioni governative e dell'IA e della sicurezza dell'IA, conviene riportare quanto affermato dalle Camere del Congresso americano, nel National Artificial Intelligence Initiative Act of 2020:

I delegati ritengono che i sistemi di intelligenza artificiale abbiano il potenziale per trasformare ogni settore dell'economia degli Stati Uniti, aumentando la produttività, migliorando la ricerca scientifica e aumentando la competitività degli USA, e che il governo degli Stati Uniti dovrebbe utilizzare questa iniziativa per favorire i vantaggi di un'intelligenza artificiale affidabile, impedendo al contempo la creazione e l'uso di sistemi di intelligenza artificiale che si

[237] https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence; https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/artificial-intelligence-european-perspective.
[238] https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai.

comportino in modo da causare danni. I delegati ritengono inoltre che questi sistemi di intelligenza artificiale dannosi possano includere sistemi ad alto rischio che mancano di sufficiente robustezza per prevenire attacchi avversari; sistemi ad alto rischio che danneggiano la privacy o la sicurezza degli utenti o del pubblico in generale; sistemi di intelligenza artificiale generale che diventano autocoscienti o incontrollabili; e sistemi di intelligenza artificiale che discriminano illegalmente contro categorie protette di persone, anche sulla base di genere, razza, età, disabilità, colore, credo, nazionalità o religione. Infine, i delegati ritengono che gli Stati Uniti debbano adottare un approccio di tutto il governo per la leadership nell'intelligenza artificiale affidabile, anche attraverso il coordinamento tra il Dipartimento della Difesa, la Comunità dell'Intelligence e le agenzie civili.

50 Il futuro dell'IA

L'intelligenza artificiale ha ancora molti limiti e molte occasioni di miglioramento. Il suo progresso finora è stato guidato principalmente dal rapido aumento nella velocità di calcolo dei computer a disposizione e anche dai progressi fatti nella progettazione degli algoritmi.

Mentre al momento attuale la tecnologia comincia a essere ampiamente disponibile, nei prossimi anni il progresso si concentrerà probabilmente su due aspetti: umano e tecnologico.

L'aspetto umano copre principalmente la capacità dell'IA di inserirsi adeguatamente nei **processi decisionali** attualmente demandati a esperti. Al momento la tecnologia dell'IA sta raggiungendo la maturazione, ma non trova ancora largo utilizzo nei processi decisionali; inoltre sono presenti casi in letteratura di un inadeguato inserimento dell'IA in tali processi[239]. Da questo punto di vista i prossimi anni vedranno la conquista di maggiore consapevolezza delle capacità dell'IA e dei suoi limiti, con il conseguente adeguamento del ruolo dell'IA nei processi decisionali. Al momento l'IA sta cercando il suo posto come "sostituto" del decisore umano, ma come dimostrato dagli esempi di errori nell'applicazione dell'IA, la tecnologia e la nostra società non sono ancora sufficientemente mature per poter ipotizzare una sostituzione totale del decisore umano. In questo senso, nei prossimi anni potremo vedere l'IA utilizzata in uno dei seguenti due modi.

Il primo modo riguarda la soluzione automatica di task semplici da automatizzare. L'IA ha dimostrato di essere in grado di poter automatizzare in maniera efficace task cognitivi semplici e ripetitivi, con la possibilità di poter migliorare l'efficienza del processo e di demandare agli esperti umani solo le decisioni riguardanti casi più complessi. Un esempio riguarda la valutazione di casi clinici, in cui si fanno esaminare dall'IA tutti i casi suddividendoli in normali e anormali, lasciando poi l'approfondimento dei casi anormali a medici esperti.

Il secondo modo riguarda l'analisi di dati e proposte di *side opinion*, ovvero l'elaborazione di dati con la stessa modalità relativa ad algoritmi intesi per compiere scelte autonome, studiando però la presentazione dei dati e delle informazioni elaborate in maniera tale da non sostituire i decisori umani, ma consentendo di ottenere un'opinione aggiuntiva, come se fosse ricevuta da un altro decisore esperto. Questa modalità, se integrata correttamente nel processo decisionale, consente di inserire l'IA nella modalità di scelta tipica dei panel di esperti, abituati al confronto, e consente di apportare un significativo aiuto nelle decisioni grazie alla capacità dell'IA di sintetizzare grosse moli di dati in un parere unico. In questa maniera l'output dell'IA può diventare un ottimo strumento nelle mani di utenti esperti. Un esempio riguarda l'utilizzo da parte di politici e governanti di una IA in grado di fornire previsioni sull'impatto di scelte politico-economiche. In questa modalità sarà necessario che gli esperti sappiano valutare gli output dell'IA, anche nel caso in cui - con il progredire degli anni - queste potranno arrivare ad aver visto una tale mole di dati da riuscire a trovare legami inintelligibili per qualunque esperto. Perciò acquisirà sempre più importanza la capacità dell'IA di spiegare le motivazioni e i processi che l'hanno portata a formulare un certo output, e anche a fornire

[239] https://time.com/5520558/artificial-intelligence-racial-gender-bias/

una indicazione di confidenza dell'output proposto, intesa come una misura della sua affidabilità.

Tuttavia, per poter arrivare a questo tipo di impiego occorre che l'IA sia affidabile in una misura adeguata al compito previsto, in maniera che eventuali errori ad alto impatto accadano molto più raramente rispetto a errori a basso impatto. Dal momento che l'affidabilità dell'IA si esprime con l'accuratezza, sarà corretto pretendere maggiore accuratezza per gli errori più impattanti.

C'è poi il problema dell'utilizzo dell'IA in quegli ambiti applicativi in cui, dopo aver considerato una mole ingente di dati, occorre richiedere **rapidità delle decisioni**. Che è proprio quello che avviene, in alcuni casi, nell'ambito della sicurezza informatica: si pensi ad esempio agli IDS, che devono bloccare tentativi di intrusione e di attacco il più velocemente possibile. In questi casi, potrebbe non esserci il tempo materiale per coinvolgere degli esseri umani nelle decisioni da prendere.

Parlando del futuro, ci si aspetta che nel corso dei prossimi anni il campo dell'IA venga influenzato fortemente dal **quantum computing**. Anche in questo caso occorre considerare sia l'effetto dell'hardware, sia quello degli algoritmi.

Nel campo dell'hardware, le tecnologie quantistiche promettono di produrre dei processori che – grazie alla possibilità di sfruttare il cosiddetto *parallelismo quantistico* e l'*entanglement* tra qubit, risolvono in tempi accettabili dei problemi per i quali i computer classici impiegherebbero tempi astronomici. L'obiettivo dichiarato è infatti quello di raggiungere la cosiddetta **supremazia quantistica**, che consiste nel dimostrare la superiorità - in termini di velocità di calcolo - dei computer quantistici rispetto a quelli classici nella risoluzione di problemi (indipendentemente dall'utilità degli stessi, cosa che naturalmente ha creato nella comunità scientifica molte discussioni). In ogni caso, l'accelerazione data dai processori quantistici su alcuni tipi di algoritmi allarga chiaramente la platea dei problemi risolvibili in pratica.

Dal punto di vista algoritmico, i concetti teorici su cui si basa il quantum computing – il fatto che un *qubit* possa trovarsi in stati che sono sovrapposizioni quantistiche di 0 e 1, il fatto che anche vettori di qubit possano esistere in sovrapposizioni quantistiche, il fatto che gli stati di due o più qubit possano trovarsi in stati entangled – costituiscono fonte di ispirazione per **nuovi algoritmi non classici**, come ad esempio nuovi classificatori[240], algoritmi di quantum machine learning[241] e, più in generale, algoritmi di IA quantistica[242], come reti neurali e algoritmi di ottimizzazione. In tutti questi casi l'IA quantistica, sfruttando anche lo sviluppo dell'hardware quantistico, permetterà l'implementazione di algoritmi di IA più efficienti, soprattutto per quanto riguarda la fase di addestramento e le operazioni di ottimizzazione necessarie per realizzare modelli efficaci. Inoltre, lo sfruttamento dei fenomeni quantistici consentirà di ottenere elaborazioni concettualmente diverse, non possibili con gli attuali computer e algoritmi classici.

Andando ancora più in là nel tempo, uno dei sogni dell'IA è quello di realizzare una **artificial general intelligence (AGI)**, a volte chiamata anche **IA forte**. Si tratta di una IA che è in grado

[240] https://ibm-q4ai.mybluemix.net/
[241] https://www.tensorflow.org/quantum/concepts?hl=en
[242] https://research.google/teams/applied-science/quantum/

di fare tutto quello che può fare un essere umano. Il fatto che un tale obiettivo sia realizzabile, anche a lungo termine, è una questione piuttosto dibattuta all'interno della comunità scientifica. I ricercatori seri che si occupano della questione sono in numero limitato, ma sono concordi sul fatto che, se la costruzione di una AGI o IA forte è possibile, allora bisogna ispirarsi al funzionamento del cervello umano, lasciando alle stesse macchine – attraverso meccanismi evolutivi – il compito di progettarne i dettagli (compito che potrebbe essere al di fuori della portata degli esseri umani, a causa della sua complessità).

Alcuni ricercatori mettono in guardia a proposito della cosiddetta *singolarità tecnologica*, un fenomeno che si innescherebbe nel momento in cui una IA dovesse raggiungere il livello di intelligenza di un essere umano: a causa dei soli progressi hardware, per non parlare di quelli algoritmici, tale IA inizierebbe a inventare nuovi hardware e nuovi algoritmi, che produrrebbero IA ancora più intelligenti, sempre più velocemente. A quel punto gli esseri umani vivrebbero in un mondo che non capirebbero più, perché per loro troppo complicato; ogni decisione verrebbe necessariamente presa da una IA, e il futuro dell'uomo in un mondo simile sarebbe alquanto incerto. C'è da dire che nel mondo scientifico ci sono molte critiche a questo tipo di speculazioni, che alcuni bollano come fantasiose, buone solamente per racconti e film di fantascienza. Viene infatti evidenziato il fatto che gli argomenti posti a favore della possibilità di simulare un cervello umano si scontrano con le attuali conoscenze sul funzionamento del cervello, ritenute ancora troppo approssimative; inoltre, diversi neuroscienziati sono del parere che le conoscenze necessarie per simulare un cervello umano non saranno disponibili neanche tra parecchi decenni.

Insomma, da un lato una AGI potrebbe essere molto utile per la risoluzione dei problemi perché costituirebbe una sorta di "coltellino svizzero" multiuso. D'altra parte, la realizzazione di una AGI potrebbe rivelarsi un boomerang per l'esistenza stessa dell'umanità. Quello che è certo è che la realizzazione di tale tecnologia è impossibile con le conoscenze tecniche e scientifiche attuali, ma naturalmente la questione non può essere semplicemente ignorata.

https://www.youtube.com/watch?v=7Pq-S557XQU

"Humans Need not Apply "

Il video di CGP Grey del 2014 che per primo ha analizzato il problema della possibile perdita di lavoro data dall'arrivo dell'IA. E' ancora oggi uno dei migliori video sul tema.

(durata: 15 minuti)

https://www.youtube.com/watch?v=3TYT1QfdfsM

"The stop button problem "

Tutti i problemi legati a come fermare un'IA forte, anche solo per poterne correggere eventuali comportamenti errati o farne il debugging.

(durata: 19 minuti e 59 secondi)

https://www.youtube.com/c/RobertMilesAI/playlists

"The stop button problem "

Rob Miles, l'autore del filmato precedente, ha un canale youtube suo, interamente dedicato ai problemi di sicurezza ed etici legati a eventuali e futuristiche intelligenze artificiali forti..

(durata complessiva di 4 playlist: più di 11 ore)

- OTTAVA PARTE: RISORSE

51 Standard e best practices internazionali

Solitamente, quando si parla di best practices nel mondo ICT, si fa riferimento agli standard ISO e ETSI ma al momento della scrittura di questo libro sull'intelligenza artificiale non ne sono stati ancora pubblicati.

Riportiamo l'elenco degli standard che ad oggi sono in fase di sviluppo (Draft) e relativi all'IA:

- ISO/IEC 22989 Artificial intelligence concepts and terminology;
- ISO/IEC 23053 Framework for artificial intelligence systems using machine learning;
- ISO/IEC 38507 Governance implications of the use of artificial intelligence by organizations;
- ISO/IEC 23894 Artificial intelligence – Risk management;
- ISO/IEC TR 24028 Overview of trustworthiness in artificial intelligence;
- ISO/IEC TR 24368 Artificial intelligence – Overview of ethical and societal concerns;
- ISO/IEC TR 24027 Bias in AI systems and AI aided decision making;
- ISO/IEC TR 24030 Use cases;
- ISO/IEC TR 24029-1 Assessment of the robustness of neural networks – Part 1: Overview;
- ISO/IEC TR 24372 Overview of computational approaches for AI systems.

Sono significative le raccomandazioni e linee guida relative alla protezione di dati personali usati con algoritmi di IA pubblicati da istituzioni e autorità significative:

- Report 2018 dell'autorità privacy norvegese su "IA e privacy"[243];
- Whitepaper su etica nei framework privacy per IA e big data pubblicato da IAPP (associazione internazionale leader di professionisti della privacy)[244];
- Report del dicembre 2017 del CNIL (autorità privacy francese) su etica e privacy che fornisce spunti per affrontare l'utilizzo di IA evitando di impattare sugli individui[245];
- Opinion dell'EDPS (garante europeo della privacy) sull'approccio corretto nell'uso della IA[246];
- Studio del Parlamento europeo su IA e responsabilità civile (luglio 2020)[247];
- Linee-guida in materia di intelligenza artificiale e protezione dei dati (gennaio 2019) - Convenzione 108[248].

[243] https://iapp.org/media/pdf/resource_center/ai-and-privacy.pdf.
[244] https://iapp.org/media/pdf/resource_center/BUILDING-ETHICS-INTO-PRIVACY-FRAMEWORKS-FOR-BIG-DATA-AND-AI-UN-Global-Pulse-IAPP.pdf.
[245] https://iapp.org/media/pdf/resource_center/CNIL_AI.pdf.
[246] https://edps.europa.eu/sites/edp/files/publication/20-06-19_opinion_ai_white_paper_en.pdf.
[247] https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf.
[248] https://www.garanteprivacy.it/temi/intelligenza-artificiale.

- Report dell'Agenzia dell'Unione Europea per la sicurezza cibernetica (ENISA) "*Artificial Intelligence Cybersecurity Challenges*" (15 dicembre 2020)[249];
- Il documento "Strategia italiana per l'Intelligenza Artificiale" del Ministero dello sviluppo economico (giugno 2020)[250];

Da citare infine la norma ISO/IEC 27701 "Security techniques — Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management — Requirements and guidelines". Questa norma non è specifica per IA, ma si applica anche in un contesto di utilizzo con IA e aiuta ad avere un riferimento autorevole sui controlli privacy.

Analogamente significativa ai fini della presente trattazione è la "*Risoluzione del Parlamento europeo del 20 ottobre 2020 recante raccomandazioni alla Commissione concernenti il quadro relativo agli aspetti etici dell'intelligenza artificiale, della robotica e delle tecnologie correlate*"[251].

[249] https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges/at_download/fullReport
[250] https://www.mise.gov.it/images/stories/documenti/Proposte_per_una_Strategia_italiana_AI.pdf
[251] Il documento è reperibile alla URL: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_IT.pdf

52 Libri interessanti

Autore/i	Titolo del libro in lingua originale	Titolo edizione italiana	Anno	Editore	Pagine	Il commento
Cosimo Accoto		Il mondo dato – Cinque brevi lezioni di filosofia digitale	2017	Egea	142	Breve saggio che con chiarezza introduce il lettore ai concetti di filosofia digitale indagando gli impatti tecnologici e le ricadute strategiche per lo sviluppo del business delle nuove tecnologie.
Nick Bostrom	Superintelligence. Paths, Dangers, Strategies	Superintelligenza. Tendenze, pericoli, strategie	2018	Bollati Boringhieri	522	Riflessione su come il destino della nostra specie dipenderebbe dalle azioni della superintelligenza artificiale che noi costruiremo entro questo secolo. Occorre quindi farlo in modo che non diventi ostile o poco amichevole.
Margaret A. Boden	Artificial Intelligence: A Very Short Introduction	L'intelligenza artificiale	2018	Oxford University Press (il Mulino)	184 (188)	Sono analizzate le sfide filosofiche e tecnologiche portate dall'IA, anche riflettendo sulla possibilità che i software possano mai essere intelligenti, creativi o addirittura consapevoli e mostrandoci come la ricerca dell'IA ci ha aiutati ad apprezzare come sono possibili le menti umana e animale.

Luigia Carlucci Aiello Marta Cialdea Mayer		Invito all'intelligenza artificiale	1995	Franco Angeli	150	Libro introduttivo e storico, scritto da due pioniere dell'intelligenza artificiale italiana nel 1995, quando ancora non si parlava di deep learning. Il testo è rivolto a non specialisti.
Naomi Ceder	The Quick Python Book	Python. Guida alla sintassi, alle funzionalità avanzate e all'analisi dei dati	2019	Apogeo	470	Per conoscere il principale linguaggio usato nel programmare soluzioni con intelligenza artificiale.
Clarence Chio David Freeman	Machine Learning and Security: Protecting Systems With Data and Algorithms		2018	OReilly & Associates Inc.	365	Algoritmi in Python per applicare l'IA all'analisi di malware, spam, anomalie di network, intrusioni, frodi.
Pedro Domingos	The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World	L'algoritmo definitivo. La macchina che impara da sola e il futuro del nostro mondo	2015	Basic Books Bollati Boringhieri (IT)	352 (357)	Attraverso un viaggio dentro gli algoritmi che governano Google, Amazon, Netflix e altri, l'autore accompagna il lettore verso la comprensione degli algoritmi che si programmano da soli, discutendone il significato e i possibili impatti su scienza, mercato, società e vita dei singoli.

Ian Goodfellow Yoshua Bengio Aaron Courville	Deep Learning		2017	MIT Press	775	Ian Goodfellow è l'inventore delle reti generative avversarie; Yoshua Bengio è uno dei maggiori ricercatori di reti neurali. Il libro è uno dei principali testi sulla tecnologia del deep learning, adatto a studenti ed esperti di machine learning. E' liberamente e gratuitamente consultabile anche dal sito web www.deeplearningbook.org.
Andrew Hodges	Alan Turing. Film Tie-In: The Enigma	Alan Turing. Storia di un enigma	2014	Bollati Boringhieri	762	Con la verve di una spy story, restituisce l'ambiente e il clima culturale del periodo in cui Turing è nato e si è formato, le sue brillanti idee in campo matematico e scientifico, e fa conoscere il lato umano e personale di un genio inquieto.
Eric R. Johnston Nic Harrigan Mercedes Gimeno-Segovia	Programming Quantum Computers: Essential Algorithms and Code Samples		2017	OReilly & Associates Inc.	317	Elementi di programmazione del computer quantistico, per comprendere le differenze rispetto al computer tradizionale e fornire un'introduzione allo sviluppo di modelli di intelligenza artificiale con il computer quantistico.

Jerry Kaplan		Intelligenza Artificiale – Guida al futuro prossimo	2017	Luiss	242	Kaplan, tra i maggiori esperti mondiali di IA, prevede che l'IA avrà sulle nostre vite un impatto simile a quello della rivoluzione industriale del '700 e indaga i tanti aspetti tecnologici, economici e sociali.
Garry Kasparov	Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins	Deep thinking. Dove finisce l'intelligenza artificiale, comincia la creatività umana	2018	Fandango Libri	388	Il noto campione di scacchi analizza il rapporto tra intelligenza umana e artificiale, rievoca i suoi match contro Deep Blue e insiste per un approccio lucido e meno enfatico nel rapporto fra uomo e macchina.
Alessandro Longo - Guido Scorza	Intelligenza artificiale. L'impatto sulle nostre vite, diritti e libertà		2020	Mondadori	256	L'intelligenza artificiale sta entrando con forza anche nella società italiana. Gli esperti concordano che ne deriveranno forti ricadute socio-economiche, all'interno della quarta rivoluzione industriale ora in corso; con la promessa di rilanciare la produttività, migliorare il benessere dei lavoratori ma anche il rischio di aumentare le diseguaglianze. Di certo, sta già cambiando il modo di lavorare e ancora maggiori impatti avrà sull'occupazione del futuro. Questa trasformazione pone anche sfide inedite per la tutela della nostra privacy e in generale dei nostri diritti fondamentali. Il libro si propone di aiutare a comprendere la rivoluzione in atto e le relative sfide.
Gary Marcus Ernest Davis	Rebooting AI: Building Artificial Intelligence We Can Trust		2019	Vintage	290	Per Marcus l'intelligenza artificiale demandata alle sole reti neurali ha raggiunto i suoi limiti e propone modelli alternativi, anche ibridi, che uniscano connettivismo e simbolismo.

Roberto Marmo		Algoritmi per l'intelligenza artificiale	2020	Hoepli	404	Algoritmi in linguaggio Python per analisi dati, machine learning, sistemi esperti, algoritmi genetici, reti neurali, e deep learning. Consigli operativi, metodologie, modalità di ragionamento.
Michele Mezza		Algoritmi di libertà – la Potenza del calcolo tra dominio e conflitto	2018	Michele Mezza	221	Saggio intrigante nel modo dei nuovi algoritmi che consentono di predire comportamenti futuri, controllare e manipolare le persone. Questa dimensione di subdolo condizionamento spinge l'autore a risvegliare in modo pragmatico il senso critico nel lettore.
Marvin Minsky	The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind		2007	Simon & Schuster; Illustrated edition (November 13, 2007)	400	Minsky mostra perché dovremmo espandere le nostre idee sul pensiero e come il pensiero stesso potrebbe cambiare in futuro.
Melanie Mitchell	Artificial intelligence – A guide for Thinking Humans		2019	A Pelican Book	448	Il libro offre una panoramica lucida e chiara sul panorama attuale della IA nei principali settori di ricerca e applicazione, presentandone i risultati già ottenuti e delineandone un possibile futuro.

Nils J. Nilsson	The Quest for Artificial Intelligence		2009	Cambridge University press	580	Uno studio molto completo sulla nascita della disciplina, con tutti i protagonisti, gli eventi e le scoperte. Il libro è anche disponibile gratuitamente in formato PDF pubblicato direttamente dall'autore: https://ai.stanford.edu/~nilsson/QAI/qai.pdf
Alessandro Parisi	Hands-On Artificial Intelligence for Cybersecurity		2019	Packt Publishing	342	Algoritmi in Python per applicare IA all'analisi di malware, spam, anomalie di network, intrusioni, frodi.
Lorenzo Pinna		Intelligenza artificiale – Nel futuro c'è ancora posto per noi?	2018	CentoAutori	264	In questo libro, scritto da un grande divulgatore scientifico, si ripercorrono la storia e le sfide affrontate dall'intelligenza artificiale. Ci si interroga inoltre su cosa avverrà in futuro, e su quale sarà il nostro rapporto con le macchine intelligenti.
Stefano Quintarelli		Intelligenza artificiale. Cos'è davvero, come funziona, che effetti avrà	2020	Bollati Boringhieri	144	Un libro divulgativo che spiega in maniera chiara l'intelligenza artificiale dal punto di vista degli effetti sulle persone, cosa ci si deve aspettare, in che modo influenzerà le nostre vite.
Stuart Russell Peter Norvig	Artificial Intelligence: A Modern Approach	Intelligenza artificiale. Un approccio moderno	2020	Pearson	704	E' un vero punto di riferimento per la disciplina, costantemente aggiornato per riflettere gli sviluppi del settore.

Max Tegmark	Life 3.0: Being Human in the Age of Artificial Intelligence.	Vita 3.0. Essere umani nell'era dell'intelligenza artificiale	2018	Cortina Raffaello	452	Descrive come l'intelligenza artificiale influirà su criminalità, giustizia, occupazione, società e sul senso stesso di essere umani. Dalla superintelligenza al significato dell'esistenza, alla coscienza e ai limiti ultimi che la fisica impone alla vita nel cosmo, per far risultare qualsiasi forma di superintelligenza uno strumento sempre più potente al servizio del benessere della persona umana.
Samuel Woolley	The Reality Game – How the next wave of technology will break the truth		2020	Endeavour	272	Esperto di propaganda digitale, l'autore – consulente della NATO del Congresso americano e del Parlamento britannico – passa in disamina recenti accadimenti che hanno messo in luce aspetti negativi di un cattivo utilizzo della tecnologia. Con storie di mercenari digitali, i black hat hacker e ricercatori governativi mostra invece come sia possibile un utilizzo della tecnologia al servizio delle persone promuovendo una cultura basata sull'accountability e la trasparenza.

53 Strumenti dimostrativi

Condizione necessaria alla base di qualunque progetto di machine learning è la disponibilità di un dataset di qualità che descriva, con un campione significativo, il fenomeno da analizzare e su cui modellare le funzioni analitiche.

Per venire incontro a questa necessità, istituzioni come il Canadian Institute for Cybersecurity hanno messo a disposizione decine di dataset[252] che contengono traffico di rete catalogato, malware, esempi di attacchi DDoS e botnet, solo per citarne alcuni.

Il capostipite di questa collezione è stato l'NSL-KDD, un dataset per costruire modelli per la classificazione di network intrusion detection, liberamente scaricabile, rilasciato nella sua prima release 1999 KDD Cup, contenente 125.973 record nel set di training. Fu sviluppato per il DARPA Intrusion Detection Evaluation Program dai MIT Lincoln Laboratory. Fornisce un tcpdump di traffico raw collezionato in una local area network (LAN) e contiene traffico normale ed attacchi che ricadono in quattro macro-categorie[253]:

- DOS: denial-of-service;
- R2L: accesso non autorizzato da una macchina remota;
- U2R: accesso non autorizzato a privilegi locali di superuser (root);
- Probing: sorveglianza.

Gli attacchi sono poi suddivisi in ben 22 sottoclassi. Questo dataset ha il difetto di essere sbilanciato tra traffico normale e tipologie di attacchi, ma è stato per anni un buon punto di partenza per costruire modelli di intrusion detection system.

Il più recente dataset CIC-IDS2017, con oltre 50 GB di dati, contiene invece ben 5 giorni di traffico durante i quali sono stati simulati attacchi che includono: brute force FTP, brute force SSH, DoS, Heartbleed, web attack, infiltration, botnet e DDoS.

Per quanto riguarda, invece, i framework di machine learning che possono essere impiegati per la costruzione di modelli analitici in ambito cyber security, non possiamo non citare Scikit-Learn[254], il framework open source in Python (licenza BSD e utilizzabile commercialmente), per lo sviluppo di modelli predittivi e che ingloba ulteriori librerie matematiche e di rappresentazione grafica come NumPy, SciPy e matplotlib.

Un interessante esempio di utilizzo in ambito di cyber security è rappresentato dal libro "Machine Learning and Security"[255]. In questo libro viene descritta la costruzione di network intrusion classifier con Scikit-learn e basato sul dataset citato prima, ovvero l'NSL-KDD.

---

[252] https://www.unb.ca/cic/datasets/index.html
[253] http://kdd.ics.uci.edu/databases/kddcup99/task.html
[254] https://scikit-learn.org/stable/
[255] Clarence Chio e David Freeman. *Machine Learning and Security*. USA: O'Reilly Media, Inc., 2018. https://www.oreilly.com/library/view/machine-learning-and/9781491979891/

Un altro importante framework distribuito anche in modalità open source è Anaconda[256]. Ingloba migliaia di pacchetti e librerie open-source e semplifica la gestione degli ambienti di sviluppo in Python/R per progetti in ambito data science e machine learning.

Per quanto riguarda l'analisi di minacce informatiche, nonché di connessioni sospette su registri netflow, DNS o proxy, un modello di analisi utile è Apache Spot[257]. Il tool permette di classificare comportamenti normali e anomali, trattando la raccolta di log relativi a un IP come un documento e utilizzando la *latent dirichlet allocation* (LDA) per scoprire strutture semantiche nascoste nella raccolta di tali documenti. LDA è un modello probabilistico generativo utilizzato per dati discreti, come i corpora di testo. LDA è un modello bayesiano a tre livelli in cui ogni parola di un documento viene generata da un insieme di argomenti sottostanti. Apache Spot deduce un modello probabilistico per il comportamento di rete di ogni indirizzo IP. A ciascuna voce del registro di rete viene assegnata una probabilità stimata (punteggio) dal modello. Gli eventi con punteggi più bassi vengono contrassegnati come "sospetti" per ulteriori analisi.

Utile strumento centralizzato per il monitoraggio e l'analisi della sicurezza è Apache Metron[258]. Metron fornisce funzionalità per l'aggregazione dei registri, l'indicizzazione completa dell'acquisizione dei pacchetti, l'archiviazione, l'analisi comportamentale avanzata e l'arricchimento dei dati, applicando le informazioni più recenti di intelligence sulle minacce all'interno di un'unica piattaforma.

256 https://www.anaconda.com/products/individual
257 https://spot.apache.org/.
258 http://metron.apache.org/.

54 Progetti finanziati

Negli ultimi anni la Commissione europea ha definito politiche e strategie sempre più orientate a porre l'attenzione su tematiche di cyber security e intelligenza artificiale. In particolare è stato istituito lo Shaping Europe's digital future, un approccio europeo alla trasformazione digitale volto a promuovere attività di sensibilizzazione, progetti e iniziative legate all'IA.

Sempre in ambito europeo esiste la piattaforma Horizon 2020 Framework Programme[259] che ha l'obiettivo di:

- fungere da punto centrale per raccogliere e fornire accesso a conoscenze, algoritmi e strumenti relativi all'IA;
- supportare i potenziali utenti dell'IA per facilitarne l'integrazione nelle applicazioni;
- facilitare l'interazione con i portali di dati esistenti necessari per gli algoritmi di intelligenza artificiale e le risorse, come HPC o il cloud computing, e supportare l'interoperabilità.

In Italia il MiSE ha pubblicato la strategia italiana per l'IA dove sono state definite 82 proposte di investimento nell'ambito dell'IA[260].

[259] https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/ict-26-2018-2020 e https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/ict-26-2018-2020.

[260] https://www.ansa.it/osservatorio_intelligenza_artificiale/notizie/societa/2020/07/02/intelligenza-artificiale-litalia-ha-un-piano_3322c5a5-efac-4b25-a151-2d045693f059.html.

55 Glossario

ACE (artificial conversational entity): è un programma di IA in grado di condurre una conversazione interattiva intelligente in modo testuale. (Sinonimi: chatbot, chatterbot, talkbot, livechat). [Si veda ad es.: https://youtu.be/FOMl68X2Amk].

Algoritmo: sequenza finita e ordinata di operazioni che, a partire da dati in ingresso (input), esegue elaborazioni semplici (calcoli) o complesse (automatizzazione di attività anche ripetitive), producendo un risultato (output).

Algoritmo adattivo: algoritmo che modifica il proprio comportamento e varia parte dei risultati in base a determinate condizioni, tra cui l'ambiente in cui è eseguito.
[Si veda ad es.: https://svn.python.org/projects/python/trunk/Objects/listsort.txt].

Analisi predittiva: termine ampio che descrive una varietà di tecniche statistiche e analitiche utilizzate per sviluppare modelli che prevedono eventi o comportamenti futuri. La forma dei modelli predittivi, i rischi, le condizioni e le decisioni prese variano a seconda dello score calcolato e del comportamento o evento analizzato.
[Charles Nyce. Predictive Analytics White Paper. AICPCU IIA, 2007[261]].

Apprendimento supervisionato: è un metodo di addestramento dell'IA, che include dei set di dati pre-classificati, permettendo all'intelligenza artificiale di apprendere informazioni dalle classificazioni indicate.
[Trevor Hastie, Robert Tibshirani e Jerome H. Friedman, The elements of statistical learning: data mining, inference, and prediction, Second edition]. Si veda anche il paragrafo 7.2 nel presente libro.

Apprendimento non supervisionato: è un metodo di addestramento dell'IA che non richiede set di dati pre-classificati, perché l'IA categorizza autonomamente i propri risultati. L'apprendimento non supervisionato è in grado di eseguire funzioni di apprendimento più complesse, ma potrebbe non ottimizzare, finendo per creare categorie di dati non necessarie aumentando la complessità, anziché semplificare.
[Trevor Hastie, Robert Tibshirani e Jerome H. Friedman, The elements of statistical learning: data mining, inference, and prediction, Second edition]. Si veda anche il paragrafo 7.2 nel presente libro.

Big data: termine utilizzato per descrivere i dati di elevata dimensione, tale in genere da essere superiori alle capacità degli strumenti software comunemente utilizzati. Trovano uso sempre più frequente nelle analytics, nel machine learning, e più in generale nell'intelligenza artificiale. [*IATE (Interactive Terminology for Europe)*, Big Data definition[262]].

Botnet: una rete di sistemi compromessi utilizzata da un malintezionatoper eseguire vari attacchi di rete, ad esempio di tipo denial-of-service. [Fonte: Anatomy of a botnet, Fortinet].

[261] https://www.hedgechatter.com/wp-content/uploads/2014/09/predictivemodelingwhitepaper.pdf.
[262] https://iate.europa.eu/entry/result/3551299/en-es-fr-it-la-mul.

Calcolo cognitivo: Tecnologia basata sulla ricerca relativa ai processi cognitivi umani. Studia come migliorare l'efficienza dell'IA e dei computer.

Chatbot (o Chatterbot): vedi ACE.

Data mining: è l'insieme di tecniche e metodologie che hanno per oggetto l'estrazione di informazioni utili dai big data (grandi quantità di dati) attraverso metodi automatici o semi-automatici.
[Trevor Hastie, Robert Tibshirani e Jerome H. Friedman, The elements of statistical learning: data mining, inference, and prediction, Second edition].

Deep learning (apprendimento profondo): un insieme di tecniche basate su reti neurali artificiali organizzate in diversi strati, dove ogni strato calcola i valori per quello successivo affinché l'informazione venga elaborata in maniera sempre più generale.
[Osservatori Digital Innovation, Alla scoperta del Deep Learning: significato, esempi e applicazioni, su blog.osservatori.net]. Si veda anche il paragrafo 7.4.1 nel presente libro.

GAN - Generative adversarial nets (reti antagoniste generative): due reti neurali artificiali congiunte e addestrate per migliorare l'accuratezza dei dati attraverso la competizione (gaming) reciproca.
[Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville e Yoshua Bengio, Generative Adversarial Nets, 2014].

Inverno dell'IA: periodo in cui la ricerca e i progressi tecnologici legati all'IA ristagnano. Molto spesso, un inverno dell'IA segue un periodo di alte aspettative, in cui vengono promessi agli investitori risultati non realistici ottenibili in breve tempo. La delusione che segue il non ottenimento dei risultati nei tempi previsti produce un calo degli investimenti, e
quindi il ristagno della ricerca e dei progressi.

Natural Language Processing (NLP, elaborazione del linguaggio naturale): processo di trattamento automatico mediante un calcolatore elettronico delle informazioni scritte o parlate in una lingua naturale.
[Isabella Chiari, Introduzione alla linguistica computazionale, Bari, Laterza, 2007].

Rete neurale (artificiale): modello matematico computazionale basato su neuroni artificiali ispirati alle reti neurali biologiche, usato per creare sistemi di apprendimento automatici, e risolvere problemi complessi. Si veda il paragrafo 7.4.1 nel presente libro.

Retropropagazione: algoritmo di apprendimento usato nelle reti neurali artificiali. L'algoritmo modifica i pesi della rete in modo da diminuire la differenza (errore) tra l'output prodotto dalla rete e l'output desiderato. Il nome dell'algoritmo deriva dal fatto che questa fase di modifica dei pesi inizia dal neurone di output e si proparaga all'indietro verso i neuroni di input. Si veda anche il paragrafo 7.4.1 nel presente libro.

Riconoscimento passivo: acquisizione di informazioni senza interazione diretta con il target, come avviene ad esempio nel caso di riconoscimento facciale da parte di una telecamera collegata a un sistema di IA (solitamente basato sul deep learning) e a un database di foto.

56 Autori, contributori e ringraziamenti

- Editor e team leader

Fabrizio Bulgarelli - RSM Società di Revisione e Organizzazione Contabile S.p.A. Partner - Risk and Advisory Services Leader
Cesare Gallotti - Consulente di sicurezza delle informazioni, qualità e privacy
Francesca Gatti - AUSED - Coordinatrice del GdL Osservatorio Sicurezza e Compliance
Alberto Leporati - Università degli Studi di Milano-Bicocca - Professore Associato
Federica Maria Rita Livelli - Business Continuity & Risk Management Consultant; BCI Italy Chapter Board Member (Deputy Leader)
Roberto Obialero - CLUSIT - Cybersecurity & Data Protection Advisor; CD Clusit
Luca Sambucci - Ecosistema.ai - Fondatore, consulente specializzato in applicazioni IA
Silvia Stefanelli - Studio Legale Stefanelli & Stefanelli - Avvocato
Mario Testino - ServiTecno - COO e Consigliere
Elena Vaciago - THE INNOVATION GROUP - Research Manager
Alessandro Vallega - Partners4Innovation - Coordinatore Community for Security

- Autori

Riccardo Abeti - EXP Legal - Founding Partner, specializzato in “Privacy e diritto delle nuove tecnologie"
Elena Agresti - Poste Italiane - CERT
Mauro Alovisio - Università di Torino - Avvocato; presidente Centro Studi di Informatica Giuridica di Ivrea Torino
Leonardo Antonelli - Oracle - Master Principal Sales Consultant
Orlando Arena - Consulente
Davide Ariu - Pluribus One - CEO
Stefano Barboni - Riesko - Senior Partner
Antonio Berardi - Leonardo - Cyber Security & Digital Competence Center
Gianluca Bocci - Poste Italiane - Corporate Affairs, Tutela Aziendale
Angelo Bosis - Oracle - Cloud Platform Solution Engineering Director
Fabio Bucciarelli - Lutech SpA, Lutech Group - Senior Security Advisor
Giancarlo Butti - Europrivacy - Internal Auditor
Andrea Cabras - Saras - ICT Security Expert
Andrea Caccia - ANORC - Consulente
Paola Cagliani - Eni gas e luce - Head of Big Data & Advanced Analytics
Alberto Canadè - Reply - Data Protection Officer Italy
Dario Carnelli - Codd&Date Suisse - IT Strategy & GRC Advisor
Davide Carnelli - Codd&Date - Consulente Architetture e Data Management
Andrea Castello - CSQA Certificazioni - Responsabile tecnico ISO 27001
Marco Ceccon - Lutech SpA, Lutech Group - Advisory Practice Manage
Francesco Ciclosi - Università degli Studi di Macerata - Adjunct professor
Igino Corona - Pluribus One - Chief Technology Officer
Rita Eva Cresci - IUSINTECH - Avvocato
Giuseppe Cusello - Cyber Partners - GRC Director
Corrado De Bari - Oracle - Cloud Architect - AI/ML domain specialist
Nicla Ivana Diomede - Responsabile Cybersecurity, Protezione Dati e Conformità
Ambrogio Ferretti - A2A - Senior IT Auditor
Enrico Ferretti - Protiviti - Managing Director
Sergio Fumagalli - Partners4Innovation - Responsabile Practice Data Protection
Giovanni Battista Gallus - Studio legale Array - Avvocato, ISO 27001 Lead Auditor; Fellow Centro Nexa su Internet e Società
Alice Giannini - Studio Legale Stefanelli & Stefanelli - Consultant
Mirko Gorrieri - Mead Informatica - CISO - Cyber Security Area Manager
Carlo Guastone - Sernet spa - Vicepresidente Business Development
Anna Italiano - Partners4Innovation - Avvocato - Senior Legal Consultant

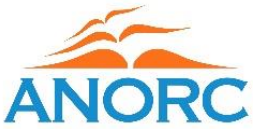
Luca Lora Lamia - KPMG Advisory - Associate Partner, Information Risk Management
Massimiliano Magri - COSTERGROUP - Smart Readiness Indicator evangelist
Davide Manconi - Eni gas e luce - Cyber Security Manager
Andrea Mariotti - EY - Associate Partner Cybersecurity & Digital Protection
Roberto Marmo - Data Scientist, Consulente e Formatore, Professore a contratto in Università di Pavia
Carlo Mauceli - Microsoft - Chief Technology Officer per Microsoft Italia
Luigi Mauro - Protiviti - Manager
Paola Meroni - Accenture - Information Security Manager
Enzo Mudu - IBM Italia - Security Technical Sales Manager
Gian Fabio Palmerini - Webuild S.p.A. - Information & Cyber Security Manager
Paolo Panza - AIT - Founder/Technical Officer
Ignazio Parrinello - Mead Informatica - Responsabile Compliance
Maurizio Pastore - Liguria Digitale - Responsabile servizi Privacy
Maria Roberta Perugini - IUSINTECH - Avvocato
Roberto Piazzolla - CLUSIT - Consulente e amministratore di sistema
Paolo Pittarello - XTN Cognitive Security - Founder e CEO
Simone Pluchino - Protiviti - Senior Consultant
Saverio Puddu - Linklaters - Managing Associate
Stefano Quintarelli - CLUSIT - Consiglio Direttivo
Giuliano Radicchi - AlgoWatt - VP Engineering
Riccardo Ranza - Consulente IT e Security
Alice Ravizza - USE-ME-D srl- Founder
Michele Rossi - SIRAM - Head of Data Science, Innovation & Hubgrade
Manuel Angelo Salvi - GRC Team - GDPR Consultant e DPO
Fabio Saulli - Cyber Partners - Principal Consultant
Aldo Sebastiani - Leonardo
Nicola Sotira - Poste Italiane - Responsabile CERT di Poste Italiane
Federico Sternini - USE-ME-D srl - Validation Engineer
Enzo Maria Tripodi - Unioncamere - Unione italiana delle camere di commercio, industria, artigianato e agricoltura - Ufficio legale e Servizio DPO

- Contributori

Luigia Carlucci Aiello , Former Professor at Dipartimento di Ingegneria Informatica automatica e Gestionale Antonio Ruberti, Sapienza Università di Roma
Francesco Bergadano , Professore Ordinario presso il Dipartimento di Informatica dell'Università degli Studi di Torino, docente di Sicurezza Informatica e Direttore del Master Universitario di primo livello in Cybersecurity
Raoul Brenna , Responsabile Security by Design e Cybersecurity Awareness, Fastweb
Alessandro Curioni , Fondatore di DI.GI Academy, specializzato in Information Security & Cybersecurity - Data Protection
Marcello Fausti , Head of Cybersecurity @italiaonline
Alessandro Longo , Direttore responsabile Agendadigitale.eu, cybersecurity360.it @Digital360. Collaboratore regolare Repubblica, Sole24ore. Autore di libro Mondadori su intelligenza artificiale
Darya Majidi , Founder & CEO di Daxo Group e di Dcare
Guido Scorza , Componente del Collegio del Garante per la protezione dei dati personali at The Italian Data Protection Authority

- Ringraziamenti

Ringraziamo le seguenti aziende ed associazioni che hanno sostenuto il nostro sforzo.



Progetto grafico a cura di Valentina Falcioni, Adriana Potoroaca e Marco Panza