

# Introduzione

L'innovazione introdotta con l'Intelligenza Artificiale (AI) è destinata a rivoluzionare le nostre vite e la nostra società. L'AI ha già iniziato a trasformare settori come quelli dell'istruzione, della finanza, dei trasporti, della difesa, dell'energia e della sanità, introducendo nuove straordinarie opportunità, ma anche molti dubbi sulle implicazioni etiche di tali cambiamenti e tante domande ancora prive di risposta riguardo le scelte che questa tecnologia ci porrà di fronte. Nel prepararci ad affrontare questo nuovo mondo, dobbiamo assicurarci che sia data a tutti la possibilità di partecipare attivamente al dibattito.

Se da un lato questi temi sono discussi da ingegneri e altri addetti ai lavori in ristretti circoli tecnici, dall'altro non c'è ancora un serio coinvolgimento del pubblico. Questa è una lacuna enorme che deve essere seriamente affrontata. L'etica dell'AI è del tutto assente nel dibattito pubblico, nel quale ancora manca completamente un'analisi comprensiva di tutte le numerose questioni etiche che vanno a intersecarsi con le tecnologie emergenti. Questo non è un compito facile, perché le sfide da affrontare sono tecnicamente complesse e intrinsecamente multidisciplinari, toc-

cando discipline quali ingegneria informatica e statistica, ma anche legge, economia, sociologia, filosofia, relazioni internazionali, e tutto ciò che vi è in mezzo.

Il mio obiettivo è riuscire a coinvolgere nel dibattito il più alto numero di persone possibile. L'importanza delle questioni ancora aperte sui rischi e l'impatto dell'AI sono tali da non poterle ignorare. Voglio fare in modo che questi argomenti siano accessibili a tutti, così da poter trovare insieme le risposte, piuttosto che permettere a poche grandi società tecnologiche (e coloro che le controllano) di decidere per noi. Anche se non mi considero un esperto su nessuno delle tante discipline che si intersecano con l'Intelligenza Artificiale, è altrettanto vero che il mio background mi consente di essere in una posizione tale da poter fare domande pertinenti e persino aiutare a trovare punti di contatto.

L'obiettivo di questo libro è di offrire ciò che manca oggi riguardo a questi temi: una panoramica completa dei rischi, delle opportunità e delle implicazioni etiche dell'ascesa delle macchine intelligenti, che organizzi in modo logico e metodico i differenti tipi di tecnologie e applicazioni. Il testo fornisce inoltre una serie di azioni concrete che cittadini, attivisti e imprese possono intraprendere per promuovere un'AI che porti benefici ma che sia anche sicura. L'intento è di avere un'analisi che risulti appetibile e interessante non solo per professionisti ed esperti, ma anche al grande pubblico, compresi coloro che esplorano argomenti così complessi per la prima volta, fornendo una miriade di casi studio ed esempi a cui tutti possano relazionarsi.

Il volume è strutturato in quattro parti.

La prima parte si concentra sulle più recenti innovazioni tecnologiche e sulle relative applicazioni. Si inizia con un'analisi dei progressi più significativi nel

campo dell'informatica, evidenziando fattori determinanti, funzionalità e i limiti tecnici. Si affronta poi una serie di questioni etiche sul ruolo sempre più rilevante svolto dalle macchine nella nostra società, con uno sguardo più da vicino a temi controversi come l'ipotesi di avere macchine in grado di raggiungere un'autocoscienza e una morale. Si esamina infine come tali tecnologie siano applicate oggi in vari funzioni e settori, evidenziandone rischi e opportunità.

La seconda parte si concentra sulle sfide etiche legate all'AI. Si inizia esaminando principi morali e questioni etiche relative all'Intelligenza Artificiale, evidenziando frequenti fraintendimenti e offrendo una carrellata dei maggiori ostacoli a un'applicazione etica dell'AI. Si fornisce poi un'analisi più approfondita in cinque casi di studio specifici: auto a guida autonoma, robot personali, armi letali autonome, realtà virtuali, chatbots e altre applicazioni basate sull'AI in grado di creare contenuti, come il caso dei chatbots.

La terza parte si concentra sui rischi più rilevanti che stiamo sperimentando con l'AI. Si inizia con un'analisi dei rischi legati alla sicurezza e all'equità dell'Intelligenza Artificiale come quelli relativi a violazioni di privacy, mancanza di trasparenza e problemi di sicurezza, per poi passare a rischi più fondamentali come la mancanza di allineamento morale, l'impatto sul mondo del lavoro, fino ad arrivare a situazioni che possano metter a rischio l'esistenza del genere umano.

La quarta parte fornisce una mappa per navigare le questioni etiche che emergono attraverso le diverse fasi del ciclo di vita di prodotti e tecnologie. Comprende, inoltre, un invito all'azione con una *roadmap* per i cittadini e gli imprenditori per favorire un'AI etica e responsabile. Le soluzioni proposte spaziano

dalla regolamentazione e dalla cooperazione internazionale, all'istruzione, alla ricerca e alla promozione della diversità. Tra gli strumenti proposti, anche un *toolkit* di *governance* per le aziende interessate a gestire attivamente la loro *pipeline* per assicurare un uso etico delle tecnologie.

L'Intelligenza Artificiale è la cosa migliore che ci sia mai successa o la peggiore? Lo scenario più probabile è che sarà entrambe le cose.

PARTE PRIMA

L'ascesa delle macchine intelligenti



## Il mondo non sarà più lo stesso

Nel 2016 il team Deep Mind di Google sfidò Lee Sedol, il diciotto volte campione del mondo imbattuto del gioco da tavolo Go. La disciplina, originaria della Cina, esiste da almeno tremila anni, è praticata da quaranta milioni di persone in tutto il mondo, ed è considerata il più antico gioco da tavolo giocato ininterrottamente fino ai giorni nostri.

Go ricorda vagamente gli scacchi, ma ha un reticolato con molte più caselle, e le sue pedine bianche e nere, chiamate pietre, hanno tutte lo stesso valore. Il gioco, tuttavia, è molto più complesso in quanto ogni turno ha circa duecento diverse mosse possibili rispetto alle sole venti disponibili con gli scacchi. Dato che a ogni turno si aggiunge una pedina, il numero di combinazioni possibili è talmente alto che è largamente superiore al totale di tutti gli atomi che esistono nell'universo osservabile. È quindi fisicamente impossibile creare istruzioni che possano aiutare a scegliere la mossa migliore in ogni scenario. L'unico modo per giocare è usare l'intuizione e la creatività, piuttosto che la memoria o formule matematiche.

Chi vinse la sfida? AlphaGo, l'AI di Deep Mind, sconfisse Lee Sedol quattro a uno. Ciò che stupì il pub-

blico non fu solo il fatto che la macchina avesse vinto contro il migliore al mondo, ma anche che lo avesse fatto con una strategia mai vista prima. Per esempio, con l'ormai famosa mossa 37 della seconda partita, sorprese tutti mettendo una pedina sulla quinta linea, qualcosa che nessun professionista avrebbe mai fatto all'inizio di un match, ma che invece si rivelò determinante per la vittoria dell'AI.

Questa non è stata certamente la prima volta che le macchine hanno vinto contro gli umani. Un famoso esempio è quando, negli anni Novanta, il computer Deep Blue di IBM sconfisse il grande maestro di scacchi Kasparov. In questo caso, al computer era stato dato un inventario di tutte le possibili mosse in modo tale da poter identificare quella migliore in ogni scenario. Come nel gioco per bambini Tris, dove ogni situazione ha sempre una mossa con più probabilità di vincita, i programmatori di Deep Blue mapparono tutte le scelte possibili e identificarono per ogni scenario la mossa vincente.

Con il gioco Go, tuttavia, ci sono troppe combinazioni per poter creare una mappa con mosse vincenti per ogni scenario possibile. I programmatori di AlphaGo si limitarono quindi a mostrare alla macchina solo una serie di mosse fatte da esperti, lasciando poi che essa giocasse contro se stessa milioni di volte, imparando autonomamente dai propri errori fino ad arrivare a inventare, in modo assolutamente creativo, nuove mosse mai viste prima nella storia del gioco.

La vittoria di AlphaGo contro Lee Sedol è stata solo l'inizio. Quello che successe subito dopo fu ancora più incredibile: una nuova versione di AlphaGo partecipò in sessanta partite contro i venti miglior giocatori del mondo vincendo ogni match con innu-

merevoli nuove mosse, mettendo in discussione tremila anni di perizia accumulata in questo antichissimo gioco. Dopo meno di un anno, Google creò poi un sistema nuovo di zecca ancora più avanzato, reso intenzionalmente ignaro delle regole del gioco Go o di esempi di giochi passati, chiamato AlphaGo Zero. In una sfida di cento partite tra AlphaGo e AlphaGo Zero, quest'ultimo vinse in tutte.

Cosa significa tutto ciò per noi?

In primo luogo, conferma che l'AI può andare oltre le capacità degli esseri umani anche nello svolgimento di attività complesse, senza neppure ricevere istruzioni. Vi immaginate un neonato che in poche ore è in grado di completare fino alla fine un qualunque videogioco? E tutto questo senza alcuna nozione di concetti base come il movimento degli oggetti o del passare del tempo, e senza alcuna indicazione su come giocare o persino le regole del gioco! Per l'AI, sapere che c'è un'immagine sullo schermo e che c'è un punteggio da battere è sufficiente per imparare a giocare a un livello superiore a quello umano. E se le macchine sono molto più brave degli umani in un gioco tanto complesso quanto Go, è ragionevole aspettarsi che la stessa cosa sarà vera come minimo anche per le tantissime attività umane meno complesse.

In un contesto in cui le macchine sono superiori all'uomo in un numero crescente di attività, che conseguenze avrà tutto questo sulla nostra società? Per esempio, che tipo di impatto avrà sul mercato del lavoro? Se le macchine sono in grado di sostituire facilmente gli esseri umani non solo in lavori ripetitivi a bassa complessità, ma anche in attività altamente complesse e creative, è chiaro che esse toccheranno tutte le professioni, comprese quelle tradizionalmen-

te considerate immuni, come quella dei medici o degli avvocati. Come si svolgerà la transizione verso questo nuovo mondo per le nostre comunità? Saremo tutti prima o poi sostituiti dalle macchine?

In secondo luogo, il caso di AlphaGo ci costringe a mettere in discussione alcune delle credenze più diffuse su ciò che differenzia il mondo organico da quello inorganico. Per esempio, dato che l'AI può eseguire attività estremamente complesse, non sorprende che si sia cominciato a mettere in discussione il ruolo della nostra intelligenza come uno dei fattori di differenziazione rispetto alle macchine. E allora cosa ci rende diversi o speciali? Ci sono alcune componenti dell'intelligenza umana che le macchine non saranno mai in grado di sviluppare? O è tempo di passare la fiaccola dell'intelligenza alle macchine che abbiamo creato, come una sorta di testimone in una staffetta della storia? Perderemo presto il diritto esclusivo di usare *sapiens* come indicatore della nostra specie? Ci stiamo davvero dirigendo verso un mondo in cui i robot prenderanno il nostro posto e gli esseri umani non avranno più alcuna rilevanza?

Più si esplora tale argomento, più la discussione diventa meno sugli algoritmi e sui robot e più su noi stessi e su cosa significa essere umani. Siamo davvero così speciali come ci consideriamo? Nonostante gli sforzi della ricerca volti a identificare le proprietà uniche del nostro cervello, non è stato ancora dimostrato che ci sia qualcosa impossibile da replicare. Se l'intelligenza umana è il risultato dell'elaborazione delle informazioni eseguita tramite reazioni chimiche nel nostro cervello, attualmente non ci sono prove dell'esistenza di un qualcosa di speciale che non possa essere duplicato da un software.

Dopotutto, per quanto ne sappiamo, il nostro cervel-

lo è costituito da un blob di particelle subatomiche che hanno esattamente le stesse proprietà delle particelle che vediamo ovunque sul nostro pianeta e in tutto l'universo. Scopriremo mai se c'è qualche aspetto non computazionale della nostra intelligenza che non potrebbe mai essere replicato da una macchina? Per il momento la questione è del tutto aperta. È necessario poi considerare che il nostro cervello ha importanti limitazioni fisiche dovute al fatto che è necessariamente confinato in un cranio sufficientemente piccolo per consentire ai nuovi individui di attraversare il canale del parto. Una volta nato, l'individuo è poi soggetto a un lungo e rischioso processo di sviluppo basato su input provenienti dall'ambiente circostante. Le macchine non hanno bisogno di tutto questo, poiché possono replicarsi rapidamente e possono crescere senza gli stessi limiti fisici.

L'accelerazione esponenziale del progresso tecnologico ci pone all'inizio di una nuova era. Di generazione in generazione crescono innovazioni tecnologiche che cambiano il mondo. L'AI è una di queste. Dopo le tre rivoluzioni industriali che hanno visto l'avvento dei motori a vapore, dell'energia elettrica e delle tecnologie digitali, siamo all'inizio di una quarta rivoluzione industriale che cambierà radicalmente la nostra civiltà e cambierà la nostra società molto più velocemente delle rivoluzioni precedenti. Abbiamo costruito macchine per aiutare gli esseri umani con il lavoro fisico, e poi computer per aiutare a fare calcoli complessi in base a istruzioni specifiche. Con le capacità di apprendimento dell'AI, abbiamo macchine che non richiedono più istruzioni, ma solo obiettivi. Il cambiamento è così grande da essere considerato la quarta rivoluzione industriale. Le opportunità sono senza precedenti, ma lo sono anche i rischi.

Da un lato l'Intelligenza Artificiale ha il potenziale per aiutare a risolvere i più grandi problemi della nostra civiltà, dalla disuguaglianza sociale, alla cura di malattie gravi, ai cambiamenti climatici. In alcune aree ci libererà dal peso di dover fare cose che non vogliamo fare, perché ripetitive, logoranti o pericolose; in altre ci permetterà di migliorare la qualità di ciò che facciamo e persino di svolgere attività che non abbiamo mai realizzato prima.

Dall'altro, aumenterà il rischio di effetti dannosi come, per esempio, la violazione della privacy, la discriminazione razziale e la manipolazione di individui o intere comunità. Alcuni esperti ritengono che l'AI stia già mettendo a serio rischio i sistemi democratici e che, a lungo termine, possa anche innescare eventi ancor più catastrofici, fino a causare l'estinzione della razza umana. Considerando le sue straordinarie capacità ma anche i rischi, alcuni hanno già iniziato a riferirsi all'AI come alla nostra invenzione finale.

Non sappiamo come e quanto velocemente l'Intelligenza Artificiale si evolverà, ma sappiamo che ci stiamo dirigendo verso un mondo dove essa sarà sempre più potente. Con l'invenzione del motore a vapore, siamo andati oltre la forza fisica umana. Le attività svolte dagli esseri umani sono quindi progressivamente passate da un lavoro prevalentemente fisico a uno basato sulle nostre capacità cognitive e decisionali. Con l'AI stiamo andando oltre le capacità di pensiero umano a un punto tale che abbiamo già iniziato a delegare alle macchine compiti altamente complessi come l'analisi di grandi banche dati e previsioni statistiche. Perfino la programmazione dell'AI stessa è stata già largamente affidata a software dotati di AI. Un'Intelligenza Artificiale che sviluppa un'altra Intelligenza Artificiale!

Siamo pronti a tutto questo? Come possiamo mitigare così tanti rischi? Come possiamo impedire che un potere senza precedenti finisca nelle mani sbagliate o fuori dal controllo umano? Cosa possiamo fare per garantire che l'AI sia utilizzata eticamente e che vada a beneficio di tutta l'umanità piuttosto che metterla in pericolo? Come possiamo continuare a stare al posto di guida e non essere colti di sorpresa dalle conseguenze impreviste dell'AI? E se si combinasse la potenza dell'AI con la forza fisica dei robot?

Due anni dopo la vittoria illuminante di AlphaGo un altro momento cruciale fu la pubblicazione, da parte di Boston Dynamics, di un video dove il loro robot-cane, Spot, riusciva con successo ad aprire una porta e andare in un'altra stanza nonostante una persona tentasse di fermarlo.

Dopo aver visto il potere dell'AI incarnato in una forma fisica familiare simile a quella di un essere vivente, le domande filosofiche sulle macchine intelligenti hanno assunto una forma molto concreta. Da un lato è difficile non essere entusiasti del suo impiego in attività pericolose come l'edilizia, l'estrazione mineraria o il monitoraggio delle sezioni radioattive delle centrali nucleari; anche l'esplorazione degli oceani e dello spazio interplanetario trarrebbe grande beneficio da questa tecnologia, consentendo alle macchine di raggiungere luoghi che nessun essere umano potrebbe sperimentare direttamente.

D'altra parte, è molto facile immaginare scenari in cui tali macchine non funzionino come previsto, sfuggano al controllo o finiscano nelle mani sbagliate. Dato l'evidente legame della tecnologia con le applicazioni militari, non è difficile credere che a un certo punto macchine simili saranno utilizzate dalle forze

di polizia o dai criminali sfruttando non solo la forza fisica, ma anche la capacità di scansionare volti, registrare conversazioni, e così via.

Vogliamo davvero essere circondati da questo tipo di macchine? Quali attività vogliamo che svolgano e quali che non svolgano mai? Considerando la potenza dell'AI accresciuta e applicata a vari sensori, telecamere, radar e la sua capacità di operare in modo indipendente e persino di prevedere le nostre mosse, come possiamo garantire che questo tipo di tecnologia sia sempre sicuro? Ci sarà sempre un grande "pulsante rosso" facile da raggiungere per disattivarla? Qual è la linea che le macchine non dovrebbero mai attraversare?

Più recentemente, con il lancio del chatbot ChatGPT di OpenAI nel 2022, si è dato accesso a centinaia di milioni di persone di un formidabile modello di AI in grado di generare testi originali con qualità comparabile a quella umana, aprendo la strada a nuove possibilità ma anche a nuovi rischi e interrogativi sul futuro della nostra società e del concetto stesso di cultura. Cosa significherà vivere in un mondo in cui testi, da romanzi a pareri legali, e tanti altri tipi di contenuti come immagini o musica, sono creati non da persone ma dall'AI? Che impatto avrà sul modo in cui impariamo, pensiamo, comunichiamo e partecipiamo al processo democratico?

In considerazione di tutto ciò, dunque, come possiamo assicurarci che la nostra civiltà userà il potere dell'AI per far sì che sia davvero la cosa migliore mai sperimentata piuttosto che la peggiore?

È ora di discuterne.

# Ti interessano questi argomenti ?

Acquista ora il LIBRO completo su

**amazon**



**Matteo Di Michele**

Intelligenza artificiale. Etica, rischi e opportunità di una tecnologia rivoluzionaria

5,0  1

