INTELLIGENZA ARTIFICIALE_ QUINTARELLI_2020_(Bollati Boringhieri)[1]

 # 1. Storia dell'Intelligenza Artificiale

La storia dell'IA inizia nell'Ottocento, con lo studio e la progettazione delle prime macchine programmabili ad opera di Charles Babbage. Quei dispositivi avrebbero dovuto permettere di elaborare diversi algoritmi, da scrivere su schede perforate, lette come istruzioni eseguibili. Il condizionale è d'obbligo, visto che a causa degli alti costi di produzione quelle macchine sarebbero state sviluppate solamente nel Ventesimo secolo. L'idea delle schede perforate non era nuova e si può far risalire al telaio di Jacquard, nei primi anni del XIX secolo, che le utilizzava per automatizzare la produzione di prodotti tessili, i cui ricami e decori venivano codificati proprio attraverso l'uso di queste schede. Di fatto l'espediente delle schede risulterà vincente: un'invenzione importante e il veicolo principale per comunicare ai dispositivi le istruzioni da eseguire.

Durante la Seconda guerra mondiale la tecnologia fa un balzo in avanti: è in questo periodo che importanti figure di scienziati si impegnano a sviluppare strumenti e tecnologie allo scopo di contrastare il nemico. Da Turing a Nash, grandi matematici si concentrano sullo studio e lo sviluppo di tecniche che permettano di sfruttare la conoscenza veicolata dai sistemi di trasmissione. Sempre in quegli anni prende vita l'idea che le macchine possano pensare.

Ricadono nell'ambito dell'Intelligenza Artificiale quei sistemi progettati dall'uomo in forma di software (ed eventualmente hardware) che agiscono nella dimensione fisica o digitale e che dato un obiettivo complesso, percepiscono il proprio ambiente attraverso l'acquisizione di dati, strutturati o meno, interpretandoli e ragionando sulla conoscenza o elaborando le informazioni derivate da questi, decidendo le migliori azioni da intraprendere per raggiungere l'obiettivo dato.

– I sistemi di Intelligenza Artificiale possono usare regole logiche o apprendere un modello numerico, e possono anche adattare il loro comportamento analizzando gli effetti che le loro azioni precedenti hanno avuto sull'ambiente.
– Come disciplina scientifica, l'Intelligenza Artificiale comprende diversi approcci e tecniche, come l'apprendimento automatico (di cui l'apprendimento profondo e l'apprendimento per rinforzo sono esempi specifici), il ragionamento meccanico (che include la pianificazione, la programmazione, la rappresentazione delle conoscenze e il ragionamento, la ricerca e l'ottimizzazione) e la robotica (che comprende il controllo, la percezione, i sensori e gli attuatori e l'integrazione di tutte le altre tecniche nei sistemi ciberfisici).

La virata più importante verso i robot inorganici probabilmente è dovuta invece a un autore famoso e prolifico come Isaac Asimov, che pone al centro di molti dei suoi romanzi proprio i robot e il loro rapporto con la società. A lui si deve l'elaborazione e il perfezionamento delle famose Tre Leggi della Robotica:
1) Nessun robot deve causare danni a un essere umano o permettere, per inazione, che un essere umano subisca danni.
2) Ogni robot deve obbedire agli ordini impartiti dagli esseri umani, a meno che questi ordini non siano in conflitto con la prima legge.
3) Ogni robot deve proteggere la propria esistenza, a condizione che tale protezione non sia in conflitto con la prima o la seconda legge.

Nel 1950, in un articolo scientifico che costituisce una pietra miliare dell'informatica, Alan Turing si chiedeva per la prima volta se una macchina potesse pensare; si rese conto che le parole «macchina» e «pensare» erano suscettibili di ambiguità interpretativa per cui riformulò la questione sviluppando un procedimento che fosse in grado di valutare quando una macchina può essere scambiata da un osservatore per un essere umano in una conversazione scritta. Si tratta del test di Turing, originariamente noto come il «gioco dell'imitazione» (imitation game). Il test si basa su un dialogo tra una persona e una controparte che l'umano non può vedere: se l'umano non riesce a stabilire se il partner del dialogo è un essere umano o un computer, secondo Turing si può affermare che il computer esibisce un comportamento intelligente.

Ma il gioco dell'imitazione non è l'unico grande contributo teorico che si deve a Turing. Meno noto al grande pubblico, ma di maggiore rilevanza per gli studiosi informatici, è il suo modello concettuale di computer, detto per l'appunto macchina di Turing, proposto nel 1936. È un modello di riferimento astratto, composto da una striscia infinita di carta che la macchina può far scorrere avanti e indietro, e sulla quale può scrivere e leggere simboli;

[1] Riassunto per "ECCOCI!" da Gigi Bacchetta; segnalazione errori: gigi.bacchetta@cgilpiemonte.it

applicando una regola presente in memoria, la macchina può scorrere avanti o indietro il nastro di un determinato numero di passi, procedere a un'istruzione successiva o arrestarsi. Si tratta della forma più semplice immaginabile di computer. Di fatto ogni computer ancora oggi potrebbe essere ridotto a una macchina di Turing.

Chatbot (dalla crasi di chat e robot). Il primo fu ELIZA, realizzato dall'informatico tedesco Joseph Weizenbaum nel 1964; il chatbot simulava una seduta psicoterapeutica utilizzando risposte basate su determinate parole chiave e precise strutture sintattiche. Sebbene non riuscì a superare il test di Turing, ELIZA fu una pietra miliare nello sviluppo dei chatbot, un punto di riferimento per la comunità scientifica, tanto che si guadagnò il nomignolo di «The Doctor».

Nel 2008 Google ha lanciato la funzione di riconoscimento vocale per la ricerca on line, seguita da Apple nel 2011, che ha lanciato un assistente virtuale – Siri – in grado di rispondere a semplici domande. Successivamente sono arrivati Alexa di Amazon e Cortana di Microsoft. Una menzione particolare tra i chatbot la merita certamente Tay di Microsoft. Il 23 marzo 2016 Microsoft rese accessibile via Twitter un bot chiamato Thinking About You, da cui l'acronimo del nome. Era progettato per imitare i modelli linguistici di una ragazza americana di 19 anni e imparare a interagire con gli utenti di Twitter, apprendendo dai messaggi che riceveva.

Il chatbot fu bersagliato da messaggi politicamente scorretti, inneggianti al nazismo e all'odio razziale; il chatbot apprese il linguaggio, emettendo a sua volta tweet dello stesso tenore. Dopo 16 ore dalla sua pubblicazione online, Tay aveva generato 96.000 tweet e Microsoft dovette sospendere il servizio. Il primo gioco che i ricercatori hanno cercato di padroneggiare attraverso l'Intelligenza Artificiale è stato quello degli scacchi, considerato all'inizio il test più estremo dei progressi del settore. Nel 1996, Deep Blue di IBM è stato il primo computer a sconfiggere un campione del mondo; si trattava di Garri Kasparov. L'algoritmo eseguito da Deep Blue utilizzava un metodo detto «di forza bruta», che analizzava milioni di sequenze prima di fare una mossa e sceglieva quella che massimizzava la probabilità di vittoria.

Tuttavia, anche se il metodo aveva permesso a Deep Blue di padroneggiare gli scacchi, non era risultato abbastanza efficace per affrontare giochi da tavolo più complessi, come ad esempio l'antico gioco cinese Go, che ha più mosse possibili del numero di atomi nell'universo. Nel 2016, i ricercatori dell'azienda DeepMind, oggi di proprietà di Google, hanno creato AlphaGo che ha battuto il campione del mondo di Go Lee Sedol 4 a 1 in una gara di cinque partite. AlphaGo ha sostituito il metodo della forza bruta di Deep Blue con un algoritmo che estrae delle correlazioni statistiche tra le azioni effettuate dai giocatori e utilizza poi queste correlazioni per predire la mossa successiva migliore.

Invece di esaminare ogni possibile combinazione, AlphaGo esamina il modo in cui gli umani hanno giocato in passato a Go, costruisce delle statistiche in cerca di correlazioni tra le modalità di gioco e usa il modello ottenuto per trovare la modalità vincente. I ricercatori di DeepMind hanno creato in seguito AlphaGo Zero, una versione migliorata del programma, che prescinde dall'osservazione dei giochi umani per estrarre i suoi modelli statistici. AlphaGo Zero ha appreso le regole di base di Go e ha imparato il gioco giocando contro se stesso innumerevoli volte. E AlphaGo Zero ha battuto il suo predecessore 100 a zero.

Il calcio d'inizio vero e proprio, che ha visto per la prima volta l'uso del termine Intelligenza Artificiale per indicare la disciplina che oggi conosciamo, è avvenuto il 31 agosto 1955 con una richiesta di finanziamento di 13.500 dollari alla Fondazione Rockefeller per un progetto dal titolo «Una proposta di progetto per una ricerca estiva a Dartmouth sull'Intelligenza Artificiale». I proponenti erano dei mostri sacri della storia dell'informatica: John McCarthy, Marvin L. Minsky, Nethaniel Rochester e Claude E. Shannon. Il sommario della proposta recitava: Proponiamo che nell'estate del 1956 al Dartmouth College di Hannover, New Hampshire, venga effettuato per due mesi e da dieci ricercatori uno studio sull'Intelligenza Artificiale. Lo studio deve procedere sulla base della congettura che ogni aspetto dell'apprendimento o qualsiasi altra caratteristica dell'intelligenza può, in linea di principio, essere descritta in modo così preciso da renderlo simulabile da parte di una macchina. Si cercherà di trovare il modo di far sì che le macchine utilizzino il linguaggio, formino astrazioni e concetti, risolvano i problemi ora riservati agli esseri umani e si migliorino.

Molti decenni e molti miliardi dopo, nonostante i passi avanti in aree specifiche, l'obiettivo appare ancora sfuggente e molti ipotizzano che non potrà mai essere raggiunto. Nel 1949 Donald Hebb scoprì che i neuroni comunicavano inviandosi scariche elettriche (oltre una determinata soglia di attivazione), che rappresentavano l'attività fisiologica di base dell'apprendimento e della memoria. Secondo questa interpretazione, quindi, un cervello non è molto difforme da un grandissimo e sofisticato computer. Sembrava questa, dunque, la strada per replicare l'intelligenza umana e costruire un cervello artificiale. Su queste basi, McCulloch e Pitts proposero un modello di neurone

artificiale, e Frank Rosenblatt, integrando i concetti di Hebb, sviluppò il «percettrone»: un semplicissimo costrutto informatico in cui l'output è determinato dalla ponderazione degli input. Il concetto era semplice, e fu fondamentale nella successiva formazione delle reti neurali.

Il «New York Times» scrisse che il percettone era «l'embrione di un computer elettronico e ci si aspetta che potrà camminare, parlare, vedere, scrivere, riprodursi ed essere consapevole della propria esistenza». Decisamente una dichiarazione impegnativa, che alzava l'asticella delle aspettative e che soprattutto prometteva di dirottare fondi della ricerca verso un genere di studi diversi da quelli iniziati a Dartmouth. Punto sul vivo, Minsky e Seymour Papert scrissero un articolo in cui criticavano la ricerca sui percettroni, sostenendone la limitatezza e paventandone l'inevitabile fallimento. Come risultato, fino agli anni ottanta circa, la ricerca in questo campo fu molto limitata. L'approccio «connettivista», teso a mimare il funzionamento del cervello umano, veniva infatti contestato dai ricercatori, che prediligevano l'approccio «logico» della formalizzazione della conoscenza.

Ciò che avrebbe successivamente risolto molte delle difficoltà ipotizzate da Minsky fu la creazione delle reti neurali. Si tratta di reti che collegano gli input dei neuroni artificiali con gli output di altri neuroni artificiali, creando complessissime reti in grado di risolvere problemi assai complicati. Nel 1959, appare per la prima volta il termine «Machine Learning» (Apprendimento della Macchina), ad opera di Arthur Samuel, informatico dell'IBM e pioniere dell'Intelligenza Artificiale, nell'articolo dal titolo Alcuni studi sul Machine Learning usando il gioco della dama. Il Machine Learning è un sistema che dà ai computer la possibilità di imparare qualche funzione senza essere esplicitamente programmato per farlo.

Samuel sviluppò una funzione matematica che calcolava un punteggio basato sulla posizione delle pedine sulla scacchiera. La funzione si proponeva di fornire una stima della probabilità di vincere per ogni mossa a partire da una data posizione. Il programma era costruito in modo da migliorare se stesso, ovvero «imparare», memorizzando il valore della funzione per ogni posizione nella quale si era già trovato. Il Machine Learning è oggi una delle principali aree di ricerca dell'Intelligenza Artificiale, assieme alla logica, alla pianificazione e all'ottimizzazione.

Nel 1969 viene prodotto Shakey, il primo robot in grado di riunire in sé mobilità autonoma, capacità di interpretare istruzioni e una certa autonomia d'azione. Shakey era programmato (in Lisp) per compiere una breve lista di azioni, come spostarsi da un luogo all'altro, accendere e spegnere le luci, aprire e chiudere le porte, salire e scendere da oggetti rigidi e spingere oggetti mobili. Se un operatore digitava sulla console il comando «spingere il blocco fuori dalla piattaforma», Shakey girava la propria telecamera fino a identificare una piattaforma e individuava una rampa per raggiungerla. Era poi in grado di spingere la rampa verso la piattaforma, farla rotolare e spingere via il blocco posto alla sua sommità. Shakey era lento e pieno di difetti, ma funzionava. Nei primi anni settanta avvennero altri due fatti particolarmente rilevanti dal punto di vista storico:

– La realizzazione concreta dell'algoritmo per l'apprendimento con reti neurali chiamato «back propagation» (descritto più oltre) e la realizzazione del principale sistema esperto dell'epoca, Mycin. L'obiettivo di Mycin era quello di formulare diagnosi di infezioni batteriche e raccomandare gli antibiotici più indicati, nel dosaggio opportuno, tenendo conto del peso del paziente.
– La creazione di un linguaggio di programmazione per l'Intelligenza Artificiale con approccio logico, il Prolog, concepito da Alain Colmerauer e dal suo gruppo di ricerca a Marsiglia nei primi anni settanta e poi sviluppato nel 1972. Il termine Prolog è una abbreviazione di programmation en logique, e deriva dal fatto che la programmazione avviene con enunciati di logica matematica del primo ordine.

Per un decennio, a partire dal 1982, il Giappone lanciò un grande programma pubblico per cercare di conquistare il Sacro Graal dell'Intelligenza Artificiale creando i «Fifth Generation Computer Systems», sistemi di computer di quinta generazione. Per «quinta generazione» si intendeva un salto paradigmatico nella storia dei computer. I computer a valvole erano considerati la prima generazione; quelli a transistor e diodi, la seconda; quelli con circuiti integrati, la terza; e quelli basati sui microprocessori, la quarta. Mentre le precedenti generazioni di computer si erano concentrate sull'aumento del numero di componenti in un singolo processore, la quinta generazione avrebbe dovuto essere significativamente più potente, facendo lavorare in parallelo un numero molto grande di processori e, come elemento fortemente distintivo, utilizzando un linguaggio di programmazione logica.

Negli anni ottanta la ricerca sull'Intelligenza Artificiale rallentò. Gli sviluppi teorici raggiunti fino a quel momento, pur notevoli, mostrarono che l'obiettivo di imitare l'intelligenza umana era assai più lontano di quanto si pensasse. Ne derivò un'ondata di delusione e di relativo abbandono. Il periodo è infatti noto come «inverno dell'Intelligenza Artificiale» e durò fino agli anni 2000, quando iniziarono a vedersi le prime applicazioni pratiche: nel 2002, iRobot,

una azienda fondata da tre membri del laboratorio di ricerca sull'Intelligenza Artificiale del MIT di Boston, lanciò sul mercato il primo Roomba, un aspirapolvere in grado di muoversi autonomamente per la casa, apprendendo la pianta dell'appartamento. Nel 2012, Google, consolidando le tecnologie in fase di sviluppo dal 2006 nel cosiddetto «deep learning» riuscì infatti ad addestrare un algoritmo per riconoscere i gattini nei video di YouTube.

 # 2. Intelligenza Artificiale. Applicazioni e tecnica

I dispositivi in generale percepiscono attraverso sensori, che trasformano ogni cosa in segnali elettrici, codificati usando i bit. Parte da qui il grande viaggio dell'Intelligenza Artificiale: da un singolo bit. All'interno del calcolatore tutto è numero. Qualsiasi informazione venga trasferita o memorizzata in un sistema digitale, viene trasformata in una sequenza di zeri e di uno, perché all'interno del computer l'informazione viene rappresentata usando il passaggio di corrente elettrica, che definisce gli uno, o la sua interruzione, da cui derivano gli zeri. La combinazione di queste due semplici modalità permette di rappresentare un'infinità di cose.

Tutto può essere rappresentato da una sequenza unica di questo genere. Ad esempio:
– la Divina Commedia, una volta salvata all'interno di un computer, viene trasformata in una sequenza di oltre quattro milioni e mezzo di bit;
– una fotografia scattata con uno smartphone è composta mediamente da oltre 25 milioni di bit;
– una e-mail di lunghezza media conta 610.000 bit.

Ad oggi, ogni 60 secondi vengono inviati 42 milioni di messaggi tramite WhatsApp o Messenger; ogni 60 secondi vengono inviate 188 milioni di e-mail. Riuscite a immaginare l'enorme quantità di dati che viene creata? E questa è solo una minima parte delle informazioni che in un breve lasso di tempo produciamo ogni giorno. La disponibilità contemporanea di una notevole potenza di calcolo, applicata a una vasta quantità di dati, ha reso l'Intelligenza Artificiale una tecnologia finalmente fruibile.

Questa tecnologia ha avuto negli anni una serie di periodi rosei, chiamati «estati», costellati di grandi risultati e promesse, che alimentavano grandi aspettative; nel giro di qualche anno molte delle aspettative venivano tuttavia regolarmente deluse o non rispettate, e seguivano quindi periodi di buio, «inverni» durante i quali i fondi e la ricerca si spostavano su altri progetti, nella convinzione che l'Intelligenza Artificiale fosse solo un abbaglio.

Possiamo distinguere due diverse categorie di Intelligenza Artificiale:
– l'Intelligenza Artificiale ristretta, che descrive tutti quei sistemi progettati e utilizzati per affrontare compiti ben specifici, seppure complessi, da giocare a scacchi a guidare un veicolo autonomo.
– l'Intelligenza Artificiale generale, che indica un sistema che riesce ad adattarsi in modo autonomo e a risolvere qualsiasi compito gli venga assegnato, indipendentemente dal contesto d'inserimento.

Mentre l'Intelligenza Artificiale ristretta è già una realtà che pervade le nostre giornate e i nostri spazi, l'Intelligenza Artificiale generale è ancora un'utopia non raggiungibile, sia per limiti tecnologici che per mancanza di comprensione del funzionamento di molti dei meccanismi dei sistemi cui dovrebbe adattarsi in modo autonomo e che guidano e governano la vita. Il machine learning sfrutta esattamente questo principio: La speranza espressa con «si spera» descrive la probabilità di aver ragione ma non preclude l'evenienza di sbagliare e quindi di imparare qualcosa di nuovo. L'errore è il margine che lascia spazio al miglioramento e a una nuova conoscenza.

Gli algoritmi di apprendimento automatico sono programmati per leggere migliaia, milioni o miliardi di dati, allo scopo di derivare una funzione complessa che sia in grado di descriverli in modo da saper riconoscere scenari nuovi, generalizzando quindi a situazioni sconosciute. Maggiore è la quantità di dati disponibili maggiore è in genere la capacità dell'algoritmo di generalizzare e di approssimare la funzione. Tecnicamente questa prima fase, dove l'algoritmo legge i dati, viene detta «fase di allenamento», o addestramento o apprendimento. In generale, quindi, si sceglie un modello di machine learning e lo si allena su un insieme di dati esistenti, che altro non sono che un insieme di features che descrivono un determinato dominio, come ad esempio l'insieme di pixel che rappresentano un'immagine o un insieme di valori numerici che rappresentano i risultati di un esame del sangue.

Ci sono tre tipologie di apprendimento:
– supervisionato
– non supervisionato
– con rinforzo.

L'apprendimento supervisionato viene in genere utilizzato per compiti di classificazione o di regressione. Nel caso della classificazione bisogna associare i dati in ingresso a una o più classi che rappresentano una categoria, una classe o un evento. Per intenderci: dati un insieme di pixel che rappresentano un'immagine di numeri a una cifra, associare quest'insieme di informazioni all'eventuale cifra rappresentata nell'immagine, e riconoscere così l'informazione numerica riportata nella foto. Nel caso della regressione, invece, si associa l'insieme di dati in ingresso a un valore numerico: dati un insieme di valori in input che rappresentano la situazione metereologica, l'output potrebbe essere il corrispondente valore della temperatura. Questo tipo di apprendimento si chiama supervisionato perché nella fase di training si conoscono sia i dati in input che i dati in output che vengono utilizzati per creare l'associazione.

L'apprendimento non supervisionato viene in genere utilizzato per creare dei raggruppamenti nei dati. Anche in questo caso si può parlare di classi, o cluster, ma non c'è un valore in output conosciuto e per questo l'approccio si dice non supervisionato. I dati in genere vengono raggruppati verificandone la similarità; dati simili vengono considerati parte dello stesso gruppo e vengono quindi inseriti nella stessa classe.

Nell'apprendimento con rinforzo, invece, l'agente esplora l'ambiente, materiale o immateriale, nel quale dovrà lavorare, e riceve una ricompensa quando l'azione svolta porta a un miglioramento. Riceve invece una ricompensa negativa quando compie un errore. In genere il compito dell'agente è quello di massimizzare o minimizzare il valore della ricompensa, e così facendo impara a muoversi correttamente in un ambiente sconosciuto, apprendendo quali sono le azioni corrette da compiere e quali quelle da evitare.

Ma attenzione: tecnicamente vengono spesso utilizzati termini che riconducono a comportamenti o azioni umane. Questa nomenclatura potrebbe far pensare a livelli di Intelligenza Artificiale che riconducono ad agenti senzienti. È necessario ricordare che quando si utilizzano termini come «allenare», «imparare» o «ricompensa» non si deve pensare a un agente cosciente, quanto piuttosto a un utilizzo improprio di termini a misura d'uomo applicati a tecnologie non coscienti. Le macchine non «imparano» nell'accezione usuale del termine, non vengono «allenate» e non «capiscono». Sono semplicemente algoritmi che vengono eseguiti e producono risultati. L'idea di base è semplice: gli attributi in input contribuiscono in modo diverso a costituire l'informazione.

Nel neurone artificiale, o perceptron – come lo aveva chiamato Rosenblatt –, il tutto viene implementato utilizzando delle somme pesate: a ogni canale che acquisisce un dato in ingresso viene associato un peso, un valore numerico che rappresenta quanto è importante quel particolare attributo nel calcolo del valore finale. Proviamo ad addentrarci un po' nel funzionamento. Durante la fase di allenamento supervisionato vengono letti i dati e calcolato il corrispondente valore in output, che viene confrontato con il valore reale: la differenza tra i due valori corrisponde all'errore commesso dal perceptron. Il risultato viene quindi utilizzato per rimodulare i pesi associati agli input. L'uso combinato di neuroni artificiali permette infatti di progettare complicati algoritmi in grado di inferire informazione da una grande quantità di dati.

Nella loro forma più semplice, le reti neurali sono composte da diversi livelli di neuroni. A ogni livello, ogni neurone riceve in ingresso le informazioni da tutti i neuroni del livello precedente, le processa utilizzando lo stesso algoritmo descritto per il perceptron e le dà in pasto al livello successivo. Nella fase di training, il valore in output viene confrontato con il valore reale utile a calcolare l'errore commesso dalla rete. Per anni è rimasta aperta la questione di come rimodulare i pesi degli input, visto che nelle reti neurali multistrato il problema è ricorsivo per ogni livello. La cosa fu risolta negli anni settanta dal giovane matematico finlandese Seppo Linnainmaa, che descrisse nella sua tesi di laurea come propagare all'indietro – dalle connessioni dei neuroni nel livello di output, giù fino alle connessioni dei neuroni in input – la correzione dei pesi associati a ogni singola connessione.

Questo algoritmo, denominato appunto di backpropagation, è oggi alla base dell'allenamento di reti molto complesse, composte anche da diverse centinaia di livelli e svariati milioni, o miliardi, di neuroni artificiali. Si comprende quindi come la potenza di calcolo, oltre alla quantità di dati disponibile, diventi fondamentale per poter utilizzare questi potenti algoritmi. La rete neurale viene quindi allenata a riconoscere per ogni input disponibile una particolare caratteristica o una classe di appartenenza.

Al termine di questa fase la rete è in grado di assegnare una probabilità di appartenenza a una classe per ogni nuovo input analizzato. Questi strumenti raggiungono performance impressionanti in alcuni scenari, spesso superiori a quelle che potrebbe dare un esperto umano in un campo specifico. Bisogna però ricordare che questi strumenti, sebbene particolarmente accurati, commettono ancora errori che sollevano in alcuni contesti importanti questioni etiche e morali sulla responsabilità legata alle decisioni che vengono prese, come vedremo meglio in seguito

affrontando questioni delicate come la responsabilità legata a un errore diagnostico o operativo, o la progettazione di sistemi che includono pregiudizi.

Uno dei test principali di queste tecnologie è stato per anni ImageNet. Con un enorme dataset composto da più di 14 milioni di immagini divise in diverse categorie (dai funghi ai satelliti), le immagini del dataset vengono tuttora usate per allenare le grandi reti neurali a riconoscere i contenuti. Nell'analizzare le immagini, un individuo umano commette mediamente il 5 per cento di errori; ogni 100 immagini, quindi, un uomo mediamente sbaglia a identificare il contenuto di un'immagine per cinque volte. Questi risultati erano imbattibili fino a qualche anno fa e questa attività umana era tra quelle considerate molto difficili da battere, fino a quando la capacità computazionale ha permesso di allenare reti neurali particolarmente complesse. È il caso della rete neurale progettata da un team di ricercatori di Microsoft, che nel 2015 è stata in grado di riconoscere oggetti con performance superiori a quelle di un essere umano.

Compaiono i recurrent neural networks e una loro variante, chiamata LSTM, proposta alla fine degli anni novanta da un gruppo di ricercatori tedeschi. Queste nuove strutture di rete neurale sono in grado di mantenere memoria delle informazioni ricevute e prendere quindi in considerazione nella fase di apprendimento la dimensione temporale. La rete è quindi in grado di tenere traccia del significato intrinseco di una frase e formulare di conseguenza risposte di senso compiuto. Negli ultimi anni si sono fatte strada nuove forme di reti neurali, la cui struttura è in grado di imparare gli stili di pittori, di musicisti e di molte altre forme di arte. Questa particolare tecnologia si chiama Generative Adversarial Network (GAN) e si basa su un meccanismo molto semplice da descrivere quanto complesso da mettere in atto.

Si tratta di sistemi in genere composti da due reti neurali che lavorano in simbiosi. Invece che utilizzare esclusivamente dati reali, il sistema si basa su un procedimento generativo. Una rete, chiamata il discriminante, è deputata a riconoscere se un particolare input rappresenta un originale o un falso, e viene ricompensata quando indovina. La seconda rete neurale, invece, ha come obiettivo quello di generare dei falsi, che devono essere sottoposti al discriminante con lo scopo di trarlo in inganno. Questa seconda rete viene ricompensata quando riesce nell'intento di imbrogliare la prima. La «ricompensa» non deve naturalmente essere intesa alla lettera; si tratta in effetti di un incremento di una funzione matematica che rappresenta una sorta di punteggio che la rete deve riuscire a massimizzare. La simbiosi risulta quindi in una partita tra due avversari che devono avere la meglio l'uno sull'altro.

All'inizio della fase di allenamento, la rete falsaria produce delle informazioni che sono chiaramente intercettabili e riconoscibili come fasulle, ma mano a mano che l'allenamento avanza diventa sempre più brava nella generazione delle informazioni, tanto da riuscire spesso a ingannare la controparte. Nella fase finale, la qualità delle informazioni fasulle generate è quasi del tutto identica a quella delle informazioni reali. Supponiamo quindi che la rete discriminante debba riconoscere se l'artista che ha prodotto un particolare quadro, rappresentato da un'immagine in input, sia effettivamente Van Gogh: la rete falsaria imparerà a generare immagini il cui stile sia il più vicino possibile a quello dell'artista.

La prima sfida su lunga distanza per veicoli interamente autonomi si è tenuta nel 2004 nel deserto del Mojave. Il DARPA Grand Challenge era un progetto finanziato dal Dipartimento della Difesa degli Stati Uniti; le squadre partecipanti dovevano costruire veicoli autonomi in grado di attraversare 150 km di deserto in completa autonomia. La prima edizione non vide arrivare al traguardo nessun veicolo, ma l'edizione successiva portò al traguardo 5 delle 23 squadre. Il veicolo che vinse l'edizione, «Stanley», era stato progettato dalla Stanford University, in un team capitanato da Sebastian Thrun, ora a capo di Waymo, il progetto di auto autonome di Google.

Per le auto autonome, va notato che esistono diversi livelli di autonomia. I livelli sono identificati e descritti dalla SAE International, l'ente internazionale che regola e norma l'industria aereospaziale e automobilistica. La SAE descrive sei livelli: dal livello zero, dove tutte le attività di controllo e guida sono completamente a carico dell'individuo che sta pilotando, al livello 5, dove il veicolo è guidato interamente da un agente artificiale che si preoccupa di verificare tutte le condizioni al contorno e di prendere decisioni autonome in caso ve ne sia la necessità. Le attuali auto autonome sono equipaggiate con dispositivi GPS, una serie di videocamere e soprattutto una serie di strumenti che permettono di percepire il mondo circostante in tre dimensioni. Si tratta di radar particolari, chiamati LiDAR. Auto autonome di livello 4 sono già disponibili e in alcuni Stati già circolano liberamente. La sfida che molte case stanno cercando di vincere è la produzione di auto di livello 5.

Gli ostacoli da superare sono però ancora molti. Ad esempio la quantità di dati prodotti da un veicolo autonomo di livello 5 è enorme, pari a circa 5 TB all'ora, una quantità enorme da gestire ed elaborare, per la quale serve una

grande capacità computazionale e tecnologie di Intelligenza Artificiale molto avanzate. Inoltre, l'imprevedibilità dell'ambiente è tale che il veicolo deve essere in grado di prendere decisioni autonome talvolta non prese in considerazione dai progettisti: nel marzo 2018 un veicolo autonomo, durante un test di Uber, ha investito e ucciso un pedone che stava attraversando un tratto di strada poco illuminato. Il veicolo aveva rilevato un ostacolo ma aveva reputato posizione e movimento come non pericolose, scartando l'ipotesi che l'oggetto potesse essere una persona. Purtroppo, il numero di incidenti che coinvolgono auto autonome è in aumento, spesso tuttavia a causa dell'incuria dei passeggeri, che dovrebbero vigilare sull'operato della macchina.

 # 3. Grandi temi

Cercando negli archivi online dei giornali dell'epoca si scopre che al tempo vi fu un dibattito acceso sull'uso delle calcolatrici. Si temeva che l'uso scolastico di quelle macchine avrebbe fatto perdere parte delle facoltà mentali agli studenti, che sarebbero stati un po' meno intelligenti, perché l'attività era caratteristica del «sapere» che una buona formazione deve assicurare, e così via. Di certo, oggi nessuno di noi si sognerebbe di mettersi a estrarre radici quadrate con carta e penna, ma nessuno pensa per questo che siamo diventati meno intelligenti. Quella che prima era una attività caratteristica dell'intelligenza umana se ne è tranquillamente uscita dal perimetro di ciò che consideriamo intelligente; ora la nostra intelligenza può dedicarsi a funzioni cognitive di livello superiore. Siamo davanti a un modo nuovo di produrre software, in grado di fare ciò che in precedenza non era possibile con i tradizionali algoritmi e le classiche procedure deterministiche.

Salvo rari casi – alienanti – un lavoro è composto da molte attività, che per la quasi totalità non hanno a che vedere con funzioni ripetitive di percezione e classificazione (e quindi predizione). Vi sono lavori per cui le funzioni che coinvolgono percezione e classificazione costituiscono tuttavia una parte rilevante del tempo. Lì vi saranno i maggiori guadagni di efficienza tramite l'Intelligenza Artificiale. Se un agente immobiliare passa mezza giornata a fare stime, potrà passare quella mezza giornata a contatto con i clienti, perché quella funzione sarà assistita e aumentata da una Intelligenza Artificiale con prestazioni… da calcolatrice: assai più veloce di quanto possa essere una persona. L'efficienza porta con sé aumento della produttività e quindi, a parità di volumi trattati, minor necessità di personale.

I computer venivano definiti fast idiots, idioti veloci, dato che sapevano fare solo calcoli ed eseguire procedure algoritmiche, come le ricette da cucina, ma a una velocità altissima. Progressivamente, mano a mano che si diffonderanno applicazioni software che utilizzano tecniche di Intelligenza Artificiale, il loro utilizzo si estenderà a funzioni di percezione, classificazione e predizione che in precedenza erano alla portata solo di un'intelligenza umana. Ed eseguiranno queste funzioni a una velocità incomparabilmente maggiore. Ma a una Intelligenza Artificiale programmata per eseguire un compito non si può chiedere nulla di diverso. A un sistema di Intelligenza Artificiale che sa identificare con grande accuratezza cellule tumorali non si può chiedere quanto faccia 1+2, perché non saprà farlo.

Si racconta un aforisma attribuito a Stalin: durante la guerra della Russia contro la Germania nazista avrebbe affermato, rispetto al proprio contingente di carri armati – superiori per numero ma inferiori per qualità rispetto a quelli tedeschi – che «la quantità è una qualità in sé». In effetti, sovente la scala sposta il nocciolo della questione su piani diversi. Con l'avvento della digitalizzazione, forte delle caratteristiche ricordate sopra, ecco che si annulla il divario tra la comunicazione individuale (o di gruppo ristretto) e la comunicazione di massa; in questo mondo qualunque individuo può raggiungere a un costo marginale nullo e in tempo reale platee persino più ampie di quelle raggiungibili da un editore, e il sistema regolamentare, che era disegnato per un modo materiale, si trova messo alle strette per effetto del cambiamento di scala sia della velocità sia della quantità.

Oggi, grazie a Internet e all'Intelligenza Artificiale, i messaggi vengono invece personalizzati su base individuale. Cessa di esistere il medium di massa per come lo abbiamo conosciuto e nasce un medium di massa personalizzato. Che non è un ossimoro: la comunicazione è nel contempo di massa e, grazie all'Intelligenza Artificiale, personalizzata per il singolo individuo. È bene evidenziare un punto: sebbene si dica spesso che Internet è uno strumento di disintermediazione, in realtà non è così. Internet è uno strumento di reintermediazione con caratteristiche diverse (quelle ricordate sopra). L'intermediario è il social network con il suo sistema di Intelligenza Artificiale, invisibile ma sempre presente, che gestisce la diffusione dei contenuti e le interazioni delle persone.

Non si può parlare di grandi temi dell'Intelligenza Artificiale senza parlare delle cosiddette LAWS: Lethal Autonomous Weapon Systems, ovvero le armi autonome. Se un sistema di Intelligenza Artificiale è in grado di percepire e fare predizioni, può anche essere usato per manovrare autonomamente un'arma e colpire un bersaglio.

Ancora di più: se è in grado di percepire e classificare, potrebbe anche determinare in modo autonomo il bersaglio. L'esistenza di sistemi di offesa letali autonomi permette di fare a meno dell'operatore umano nel teatro operativo, ponendo i presupposti di una trasformazione nella struttura stessa delle operazioni militari, qualitativamente diversa da precedenti innovazioni tecnologiche in ambito bellico.

Esiste un appello per mettere al bando questo tipo di armi, noto come «Stop killer robots». Il sito della campagna spiega i motivi dell'allarme: armi completamente autonome possono decidere chi vive e chi muore, senza ulteriori interventi umani, varcando così una soglia morale. In quanto macchine, queste armi mancano naturalmente di caratteristiche intrinsecamente umane, come la compassione, che sono necessarie per compiere scelte etiche complesse. Stati Uniti, Cina, Israele, Corea del Sud, Russia e Regno Unito stanno sviluppando sistemi d'arma con una significativa autonomia nelle funzioni critiche di selezione e di attacco dei bersagli. Il mondo potrebbe entrare in una destabilizzante corsa agli armamenti robotici. Sostituire le truppe con le macchine potrebbe rendere più facile la decisione di andare in guerra e conseguentemente spostare ulteriormente il peso del conflitto sui civili.

Leonardo è una multinazionale italiana che opera anche nella produzione di armamenti. Finanzia una fondazione, entità autonoma, che a novembre del 2019 ha presentato la sua proposta di Statuto etico e giuridico dell'Intelligenza Artificiale. Riguardo alle armi autonome, nel documento si leggono i seguenti requisiti:
• La finalità difensiva dell'esercizio dei sistemi autonomi.
• Assicurare il loro controllo e la validazione finale della decisione esecutiva all'uomo (controllo).
• Progettarli in modo che abbiano sistemi di tracciamento che garantiscano l'attribuzione delle responsabilità nel loro uso (responsabilità).
• I loro sistemi di apprendimento e adattamento devono documentare ed essere in grado di spiegare in modo comprensibile all'operatore umano le loro intenzioni (trasparenza e spiegabilità).
• Sulla scorta del precedente, si deve fare in modo che l'operatore umano preveda il comportamento delle loro funzioni autonome (fiducia), tanto in ambito direzionale quanto in quello operativo.
• Bisogna sviluppare codici etici professionali e addestrare operatori umani che siano responsabili del loro uso e siano chiaramente identificabili (formazione).

Sarebbe bello vivere in un mondo che non ha bisogno di armi. Vivere in un mondo senza armi autonome, invece, è necessario. I dati che vengono utilizzati per addestrare un sistema portano con sé, ben nascoste, tutte le sfumature e i pregiudizi della società che descrivono. Il mondo è diverso da come pensiamo che sia o da come idealmente vorremmo che fosse. I sociologi conoscono bene questo tema e conducono da tempo esperimenti sociali per studiare le inclinazioni della società. In uno di questi, una stessa bambina, posta sola e piangente in una via piena di negozi attira molte persone, che si offrono di aiutarla, ma solo se vestita di tutto punto. Quando è logora e malvestita, nella stessa via e nelle stesse ore, viene invece aggirata dalla quasi totalità dei passanti.

L'Intelligenza Artificiale, in pratica, non si comporta in modo etico. Non si comporta nemmeno in modo non etico. Non ha proprio idea di cosa sia l'etica. Ma noi che osserviamo le sue predizioni possiamo valutare se i risultati sono allineati o contrastanti con i nostri principi etici. Un capitolo a sé meriterebbe il trattamento dei dati di addestramento errati. Se non si controlla il processo di formazione e ottenimento del dato, non si può essere certi che esso sia corretto e possa essere usato per trovare correlazioni utili. Tra gli addetti ai lavori si usa l'espressione «garbage in - garbage out», ovvero «immondizia in entrata produce immondizia in uscita». Solo che in questo caso l'immondizia può essere molto ben nascosta e difficile da trovare (talvolta è impossibile essere certi che un dato in ingresso sia giusto o sbagliato).

 # 4. Etica dell'Intelligenza Artificiale

L'etica si interroga sulla vita e sull'agire in relazione a un ideale di bene o di felicità. Tutte le volte che ci chiediamo se un'azione sia buona nei confronti di qualcuno o possa portare felicità stiamo ragionando eticamente. Naturalmente i grattacapi sono dietro l'angolo, ed è spesso difficile capire se una particolare scelta sia buona o giusta, sia perché ciascuno può avere idee diverse al riguardo sia perché le azioni toccano quasi sempre più persone con effetti differenti. Ciò che conta in questa sede, però, non è stabilire cosa sia il bene o cosa sia la felicità, ma capire che l'etica è la disciplina che riflette su queste domande. Gli algoritmi di Intelligenza Artificiale nascono come strumenti che realizziamo per raggiungere un obiettivo di nostro interesse nel modo più efficiente possibile. Grazie all'Intelligenza Artificiale è possibile automatizzare compiti e delegare a software o macchine lo svolgimento di determinate funzioni.

La volontà di automatizzare e delegare compiti è il primo motore dello sviluppo dell'Intelligenza Artificiale. Grazie alla possibilità di progettare sistemi in grado non solo di elaborare rappresentazioni strutturate sia dei dati che dello scopo da raggiungere, ma anche di modificare il proprio funzionamento in base alle analisi dei processi che svolgono, possiamo ora contare su strumenti capaci di funzionare da sé; liberandoci dall'incomodo di dedicare tutte le nostre energie e la nostra attenzione al compito ad essi delegato. Ad assumere un'importanza nodale è dunque la questione dei valori.

La questione dei valori sorge quando i vari attori coinvolti a diverso titolo – utenti, designers, programmatori, produttori, regolatori ecc. – si rivolgono alle tecnologie di Intelligenza Artificiale o vengono coinvolti nel loro funzionamento. In questi casi i diversi attori in gioco avranno diverse convinzioni e aspettative rispetto alle finalità e alle intenzioni che è bene perseguire.

Computer Ethics: disciplina che comprende riflessioni sulla responsabilità di programmatori e ingegneri informatici, analisi dell'impatto etico-sociale delle tecnologie informatiche, raccomandazioni (o policy) per il loro uso eticamente accettabile.
Roboetica: disciplina volta a promuovere la produzione di tecnologie robotiche benefiche e ad allineare l'ingegneria robotica a standard morali e sociali. Un'etica pensata per i ricercatori, i produttori, gli utenti e gli altri attori umani coinvolti nel mondo della robotica.
Etica delle macchine: tentativo di dotare sistemi tecnologici della capacità di gestire in autonomia situazioni morali in modo soddisfacente e discussione dei relativi impatti etici, sociali e filosofici.
Etica dei dati: disciplina che si occupa di identificare e discutere varie questioni etiche collegate alla raccolta di dati e alla loro analisi attraverso sistemi di machine learning. I problemi spaziano, ad esempio, dalla privacy ai bias e dalla proprietà dei dati alla discriminazione.

In generale, lo scopo del lavoro etico è sostanzialmente duplice:
– promuovere l'allineamento tra le intenzioni delle varie parti e i valori etici pertinenti all'uso in esame;
– individuare, correggere o denunciare applicazioni volte a servire intenti eticamente inaccettabili che ignorano, o violano, valori determinanti in relazione al loro ambito di funzionamento.

L'etica dell'Intelligenza Artificiale mira ad allineare gli algoritmi ai valori pertinenti, a valutare criticamente il funzionamento e l'uso dei sistemi dal punto di vista dei loro impatti morali, a sensibilizzare la comunità scientifica e l'opinione pubblica sull'importanza di assumere consapevolmente le sfide poste in essere. Affinché esista un allineamento tra tecnologie e aspettative etiche è però necessario che queste siano chiaramente definite. Un quadro valoriale il più possibile coerente e chiaro è in effetti cruciale se vogliamo che l'attività tecnologica di ricerca e sviluppo sia eticamente consapevole e venga praticata alla luce di tale consapevolezza.

Tracciare un quadro valoriale non significa gettare le basi per una procedura in grado di risolvere ogni contrasto in modo oggettivo e inappellabile. Tantomeno è onesto trasformare un quadro valoriale in una tavola dei comandamenti. In verità, simili documenti vanno intesi come mappe che tracciano una serie di linee guida utili a sollevare, discutere e affrontare il problema dell'allineamento: non per scavalcare l'atto del giudizio, quindi, ma per innescarlo e stimolarlo. La riflessione etica sull'Intelligenza Artificiale richiede i contributi congiunti di professionisti e studiosi diversi, che si sforzino di dare, per quanto possibile, unità a una molteplicità quasi infinita di istanze. L'etica dell'Intelligenza Artificiale non è riducibile né alla scrittura di documenti sui principi che ne dovrebbero guidare il corso né alle riflessioni individuali o di gruppo su situazioni concrete e casi specifici. Il suo dominio, al contrario, comprende entrambi i momenti e vive dell'incessante movimento che li connette.

Due problemi fondamentali, in cui si declina la questione dell'allineamento:
– il problema dell'uso;
– il problema della non-neutralità degli algoritmi.

Se un danno c'è stato e la causa non è un fenomeno naturale (ad esempio un terremoto) ma è connessa ad attività umane, qualcuno deve esserne responsabile e quindi assumersene la responsabilità. D'altra parte se l'idea più immediata è quella di scagionare gli utenti per rivolgere l'indice agli altri umani coinvolti – programmatori, costruttori, eccetera – non si va molto lontano: a ogni livello l'autonomia funzionale può essere indicata come causa dell'erosione del controllo e di conseguenza, limitazione della responsabilità. Sembra rimanere solo l'opzione di ricondurre la responsabilità all'unico ente del tutto in controllo del funzionamento del sistema: il sistema stesso.

Si tratta però di un'opzione paradossale. La responsabilità morale non è tra gli aspetti che possano essere significativamente delegati a sistemi di Intelligenza Artificiale, nonostante venga ugualmente coinvolta nella delegazione del compito. Sarebbe ingenuo pensare che sistemi di Intelligenza Artificiale ed esseri umani siano equiparabili, per cui il sistema può assumersi senza scarti anche l'onere morale della responsabilità. Per quanto abbia senso interrogarsi sul gap – ovvero sul vuoto di responsabilità che si apre sotto ai nostri piedi nel rapporto con i sistemi di Intelligenza Artificiale – pretendere di colmarlo additando il sistema stesso come luogo della responsabilità non equivale a risolvere il problema, ma a nasconderlo sotto il tappeto.

Rimangono due alternative percorribili:
– recuperare una qualche forma di controllo significativo sul funzionamento autonomo dei nostri strumenti di Intelligenza Artificiale;
– estendere il modo in cui pensiamo la responsabilità al di là del paradigma del controllo.

Ugualmente doppia sarà la mossa da eseguire per sciogliere il nodo operando a due livelli:
– a livello ingegneristico, recuperando o rendendo disponibile un qualche grado significativo di controllo sulla funzione svolta dall'automatismo;
– a livello concettuale, pensando la responsabilità morale anche a partire da altri elementi che non siano il mero controllo diretto degli eventi.

Un concetto importante: il controllo umano significativo. La prima strada battuta è quella nell'ambito delle ricerche sul cosiddetto meaningful human control (MHC), il cui scopo consiste nell'elaborare stratagemmi e interfacce che consentano all'utente di esercitare un controllo significativo sugli aspetti relativi all'esecuzione automatica di un compito quando essi non risultano adeguatamente gestibili dal sistema o quando riteniamo opportuno che non sia il sistema a gestirli. Tra questi aspetti figura ovviamente il lato etico dei compiti delegati agli strumenti di Intelligenza Artificiale.

L'utente finale però è solo la punta dell'iceberg. Affinché gli utenti possano esercitare un controllo significativo sul funzionamento di una tecnologia di Intelligenza Artificiale, è necessario che tutte le parti coinvolte – programmatori, sviluppatori, distributori, regolatori e produttori – si assumano la responsabilità di ciò che è o può essere riportato sotto il loro controllo in modo da mettere l'utente finale in condizione di comprendere quale sia il proprio ruolo e la propria effettiva capacità di azione.

L'importanza del MHC è testimoniata dalla ripresa che ne è stata fatta in vari ambiti della robotica e dell'Intelligenza Artificiale: dai sistemi d'arma autonomi, ai robot chirurgici, alle automobili a guida autonoma. La sfida consiste nel:
– non mettere in competizione controllo umano e autonomia funzionale;
– chiarire quali aspetti di un compito non sia problematico delegare, e quali altri aspetti al contrario è bene riservare alla decisione umana.

Proprio come la teoria del controllo umano significativo ci invita a rivedere alcuni concetti, grazie ai quali interpretiamo il rapporto che ci lega agli strumenti di Intelligenza Artificiale, così l'impasse a cui ci conduce la riflessione sulla responsabilità ci spinge a riconsiderare il modo in cui pensiamo la responsabilità stessa. Come già sottolineato, il vuoto di responsabilità si apre solo se si ritiene di essere responsabili unicamente di ciò che si può controllare in modo diretto. Siccome nessuno è realmente in grado di esercitare questo genere di controllo nel caso dell'uso di sistemi di Intelligenza Artificiale, allora sarebbe ingiusto ritenere qualcuno responsabile di eventuali conseguenze moralmente problematiche.

Senza dilungarsi troppo, una tecnologia di Intelligenza Artificiale si trova a svolgere una determinata funzione in un determinato contesto perché determinati esseri umani hanno voluto che così fosse. Ogni automatismo è il risultato di precise decisioni prese in vari momenti da diversi attori sociali. Ciò di cui abbiamo bisogno, quindi, è un'assunzione di responsabilità da parte di tutti gli attori sociali coinvolti, basata sulla presa di coscienza delle motivazioni che portano determinate tecnologie a funzionare in precisi contesti sociali. Da ciò è poi inseparabile un'educazione al sentimento di responsabilità che sappia tanto mostrare l'inconsistenza di atteggiamenti autoassolutori quanto promuovere l'allineamento dei sistemi di Intelligenza Artificiale alle relative esigenze etiche.

Della non-neutralità degli algoritmi si sente discutere sempre più. Una delle più note formulazioni dell'argomento però, non ha a che fare con complicate tecnologie digitali, ma con un artefatto assai più tradizionale: il cavalcavia. In un memorabile articolo del 1980, Do Artifacts Have Politics?, Langdon Winner mostrò come la progettazione di un artefatto tanto banale quanto inerte come un cavalcavia potesse mediare efficacemente idee di carattere etico-

sociale e contribuire con la sua mole all'affermazione (o negazione) di specifici valori. Analizzando i cavalcavia edificati nei decenni centrali del secolo scorso dall'urbanista Robert Moses lungo le strade che conducono al parco pubblico di Jones Beach, a Long Island nello Stato di New York, Winner nota come essi presentino un'altezza stranamente limitata. Il motivo alla base della scelta di costruire cavalcavia bassi, secondo lo studioso, era quello di impedire la circolazione di autobus pubblici, in modo che il parco di Jones Beach fosse irraggiungibile dai membri delle classi meno abbienti, che non potevano permettersi l'acquisto di un'automobile, e diventasse quindi un'enclave riservata alla fascia benestante della popolazione.

L'esempio dei cavalcavia di Moses mostra intuitivamente come alcuni artefatti incorporino dei valori e come ciò sia una conseguenza del modo stesso in cui sono fatti, portando con sé quell'insieme di scelte consapevoli e inconsapevoli che ne formano il progetto. L'idea della neutralità degli algoritmi, basata sulla purezza formale del calcolo computazionale, non è che un miraggio da cui ci si deve liberare al più presto se si vogliono allineare le tecnologie alle nostre esigenze etiche.

Il problema della proiezione di convinzioni etiche in sede di progettazione può essere affrontato dedicando specifiche sessioni di lavoro a individuare e gestire eventuali bias, ovvero pregiudizi o credenze irrazionali eticamente rilevanti. Tuttavia il codice di un programma non è l'unico luogo in cui può verificarsi la proiezione di valori. Pregiudizi e convinzioni circa il bene e il male sono contenuti anche nei dati da cui il sistema apprende in modo automatico. I modi in cui pregiudizi eticamente rilevanti possono annidarsi in insiemi di dati sono molteplici: i dati possono semplicemente rispecchiare pregiudizi e convinzioni che effettivamente caratterizzano le opinioni di chi li ha selezionati e inseriti. Oppure, i bias possono derivare da generalizzazioni azzardate, basate su insiemi di dati incompleti, incoerenti o raccolti secondo procedure inadeguate.

Non si tratta però solo di assicurarsi che gli insiemi di dati sulla base dei quali le reti neurali vengono addestrate siano raccolti e organizzati in modo adeguato. Si tratta anche di proteggere la privacy di chi quei dati li produce – ad esempio tutti noi nell'uso quotidiano di uno smartphone – dall'ingordigia di chi se ne serve per estrarre informazioni e profitti, di vigilare che la dignità degli individui sia sempre rispettata e di pensare a modalità tramite cui redistribuire la ricchezza connessa alla monetizzazione dei dati. Sebbene fin qui abbia preso il sopravvento la preoccupazione per i possibili impatti negativi della non neutralità delle tecnologie, non bisogna perdere di vista l'altra faccia della medaglia. Così come pregiudizi e disvalori possono essere proiettati nelle tecnologie di Intelligenza Artificiale, allo stesso modo è possibile pensare tecnologie che abbiano proprio lo scopo di aiutarci a diffondere e concretizzare ciò che ci sta a cuore.

 # 5. La fine del lavoro: una notizia grossolanamente esagerata

Diversi studiosi hanno analizzato l'impatto dell'automazione sul capitale umano, concludendo in alcuni casi che la sostituzione della macchina all'uomo è un fattore determinante di crescita economica (a scapito della questione sociale), in altri invece orientandosi verso un approccio più equilibrato, dove la domanda e la peculiarità di alcuni lavori non viene meno ma, semplicemente, viene alterata. L'analisi dei dati non può essere improvvisata, né tantomeno inventata dal nulla. Essa richiede una preparazione, uno studio e un'attitudine che non sono facili da trovare nel mondo del lavoro. Soprattutto, prevede che vengano introdotte nuove figure professionali.

Nascono così lo scienziato dei dati e l'ingegnere dei dati. Sono due delle figure attualmente più ricercate (e sì, anche più pagate) dell'intero mercato del lavoro, nonché due dei lavori più interessanti e affascinanti secondo diversi studi e sondaggi. Vediamo più in dettaglio di cosa si occupano. Lo scienziato dei dati è l'analista per eccellenza, ovvero la persona che connette i puntini, prende in considerazione dati differenti, apparentemente non collegati fra loro, e identifica relazioni nascoste che possono apportare valore all'impresa e alle persone. È colui che trova risposte a domande che ancora nessuno ha posto, ma che non per questo sono meno importanti. Questo è un lavoro che si impara tanto sui libri quanto sul campo, dove la preparazione teorica e l'attitudine sperimentale ed empirica non possono prescindere l'una dall'altra.

Essere un bravo scienziato dei dati prevede infatti un insieme di competenze difficili da maturare: bisogna essere in parte ingegneri informatici, in parte statistici, un po' uomini d'affari, gran comunicatori ed esperti in domini particolari. L'ingegnere dei dati cura l'aspetto relativo alla corretta creazione, gestione, immagazzinamento e uso dei dati all'interno dei flussi e dei processi aziendali. L'ingegnere ha il compito di disegnare e mantenere un'architettura funzionante ed efficiente per consentire allo scienziato di avanzare teorie e analizzare i dati con tecniche statistiche. Chiudiamo il cerchio con altre due figure essenziali per un data team: l'analista di business intelligence, e l'analista di customer intelligence. Queste due figure rappresentano l'interfaccia pubblica dello

scienziato dei dati, gli interlocutori principali del team con il mondo esterno. Il primo traduce le conclusioni e le analisi in contesti aziendali, aiutando a capirne gli impatti sul business e delineando le corrette azioni da intraprendere per ottenere i risultati che i dati suggeriscono, ed è, di fatto, uno stratega. Il secondo, invece, è l'interfaccia del team verso il cliente finale, colui che gestisce l'interazione col pubblico e i dati che entrano ed escono da questi scambi. Il suo scopo finale è far sì che dati e decisioni vengano utilizzati per migliorare l'esperienza dell'utente e il prodotto di cui egli si serve.

Esistono tre possibili nuovi ruoli, al momento forse difficili da concepire, ma che potrebbero assumere una rilevanza fondamentale molto presto. L'«allenatore di macchine»; il «traduttore di macchine» si occuperebbe invece di spiegare i processi nascosti per svelare come le macchine sono arrivate a una certa azione o decisione. Diventerebbero i nuovi etologi delle macchine, che studiano e comprendono il comportamento degli algoritmi; ma anche i nuovi medici legali della tecnologia, capaci di identificare dove, come e quando un sistema ha commesso un errore, evidenziando le ragioni sottostanti. Infine, l'«ispettore degli algoritmi» avrebbe il compito di controllare e verificare continuamente che una macchina faccia quello per cui è stata programmata e nulla più. Sarebbe il professionista incaricato di ridurre al minimo la possibilità di conseguenze inattese e negative generate dall'utilizzo dell'Intelligenza Artificiale nella vita di tutti i giorni, e di creare un ambiente di fiducia verso le azioni compiute da macchine intelligenti.

Tre ruoli stanno recentemente emergendo per rispondere a questa esigenza, anche se è facile pensare che non saranno i soli. Il primo è quello del Chief Data Officer, o CDO. Il CDO nasce dalla necessità di centralizzare in un'unica figura capacità tecniche, legali e amministrative, e connettere il capo dell'infrastruttura tecnica (detto anche Chief Technical Officer, o CTO) con il capo del reparto analisi (il Chief Analytics Officer, o CAO). Il suo ruolo è quello di occuparsi della corretta gestione dei dati all'interno di un'azienda, cosi come gestire i processi attraverso i quali questi dati vengono creati, immagazzinati e usati. Un CDO ha quindi come obiettivo finale quello di garantire l'accesso alle giuste informazioni da parte di altre unità aziendali e di democratizzare i dati all'interno dell'azienda.
Il capo dell'Intelligenza Artificiale, o Chief Artificial Intelligence Officer (CAIO), in futuro sarà sicuramente determinante per la loro trasformazione. Come suggerisce il nome stesso, il CAIO è responsabile di tutte le attività connesse all'uso dell'Intelligenza Artificiale, tanto a livello di prodotto quanto di processo. Il suo compito fondamentale consiste nel guidare lo sviluppo di progetti di Intelligenza Artificiale, machine learning, e data science, che possano estrarre valore dai dati immagazzinati e analizzati da singoli reparti dell'azienda.

Si tratta di una figura estremamente tecnica, che trova la sua origine in un background fortemente accademico ed empirico, che sperimenta nuove tecnologie e applicazioni al solo fine di creare nuove fonti di valore per l'impresa. Il Chief Robotics Officer (o CRO) potrebbe infatti diventare la figura manageriale che si assumerà il compito di gestire tutti i processi automatici e la forza lavoro non umana all'interno dell'impresa. Esistono molti casi in cui una macchina può sostituire un essere umano, e ce ne sono molti altri in cui la macchina rende possibile lavori altrimenti non eseguibili. Se alcuni ruoli potranno venire meno, altre figure cambieranno, si adatteranno, ed emergeranno sulla base delle potenzialità che le macchine ci mettono a disposizione. Esistono infatti diverse stime che misurano l'impatto dell'Intelligenza Artificiale, alcune che teorizzano una perdita di posti di lavoro fino al 40 per cento nell'economia mondiale e altre che teorizzano una crescita economica dell'ordine di centinaia migliaia di miliardi.

Gli studi più recenti vanno in questa seconda direzione. La società di consulenza McKinsey stima che solo in Europa il valore dell'Intelligenza Artificiale si possa valutare nell'ordine di grandezza di 3000 miliardi, vantaggio che la comunità potrebbe raggiungere (o mancare) nei prossimi dieci anni. Equivarrebbe a un 20 per cento della crescita economica dell'area dell'euro. Nello specifico, l'incremento della domanda sarà guidato dalla capacità dell'Intelligenza Artificiale di diminuire le frizioni e le barriere strutturali per aumentare la connettività nelle catene di valore e facilitare una globalizzazione più rapida. In particolare si prevede un aumento della ricchezza tramite la creazione di nuove fonti di valore. Dal punto di vista della produttività, invece, sicuramente alcuni dei ruoli esistenti verranno potenziati, mentre altri, più tediosi e a scarso rendimento, verranno resi più efficienti. Questo renderà maggiore qualsiasi ritorno sugli investimenti, sia in termini di asset fisici che intangibili, e incoraggerà una nuova ondata di innovazione. Come per ogni predizione, si tratta solo di stime, utili a farci riflettere su come possiamo affrontare i possibili esiti negativi che scaturirebbero da uno scenario di automatizzazione completa, ma nulla più.

Che cosa succederà ai nostri lavori, dunque? Difficile a dirsi. Sebbene gli impatti economici siano estremamente complessi da predire, possiamo credere che la domanda di lavoro per compiti di routine si ridurrà nel breve termine, venendo rimpiazzata da una macchina. È più facile pensare a un algoritmo che coordini gli spostamenti di una flotta di autisti in una città che a un programma che assista anziani in una casa di cura.

I compiti altamente strutturati, ripetibili e molto specifici sono i primi a soffrire degli effetti dell'automazione (tipicamente, gli operai non specializzati). Lavori composti da molti compiti complessi, come per esempio il medico, più che scomparire del tutto probabilmente muteranno, sfruttando il potere dei dati per specifici scopi, per aumentare la produttività automatizzando in parte o in toto alcune funzioni ben delineate. Altri ruoli maggiormente legati alla creatività e all'intelligenza emotiva sembrano invece essere più protetti, almeno nel medio termine.

Quali sono i settori che subiranno un maggiore impatto? Sicuramente il settore manifatturiero è storicamente il primo ad essere colpito dall'automazione. Il trasporto e la logistica, così come il commercio online e la finanza, seguono a ruota. Da alcuni anni la professione dell'autista è tra le più difficili da reperire; si tratta di un settore in cui la domanda supera l'offerta di lavoro. Esiste una correlazione tra la crescita dei nuovi modelli di commercio elettronico e il tipo di professione; un esempio di come lo sviluppo tecnologico faccia aumentare la domanda di lavoro qualificato, ma anche di particolari tipologie di lavoro meno qualificato, con un impatto positivo sull'occupazione.

Il 2,5 per cento della forza lavoro statunitense – il settore della guida – può dormire sonni tranquilli ancora per un bel po'. Non perdiamo però di vista l'automazione nei centri logistici e la pressione su quantità, qualità e produttività del lavoro nei magazzini sorvegliati dall'Intelligenza Artificiale. I maggiori impatti sono lì, anche a causa di una riduzione delle competenze richieste e quindi del grado di sostituibilità del personale. La forza che può contrastare questa pressione è l'aumento della formazione, che può mettere le persone in grado di compiere funzioni diverse, di svolgere attività meno ripetitive, per le quali sia richiesta la capacità di affrontare e gestire eccezioni e criticità. Dovremo cioè procedere a una formazione diversa e permanente dei lavoratori per renderli in grado di affrontare la rivoluzione digitale. Non è facile, perché, a differenza delle rivoluzioni precedenti, quella digitale sta avvenendo nell'arco di una sola generazione, mettendo a dura prova la capacità di adeguamento delle strutture sociali e produttive.

Una delle grandi preoccupazioni è legata alla polarizzazione degli stipendi. Visto quanto scritto fin qui, è facile immaginare uno scenario in cui pochi lavoratori, estremamente specializzati e competenti, vengono strapagati mentre la maggior parte delle persone vive con uno stipendio di sussistenza. Se si pensa che già oggi alcuni ricercatori di Intelligenza Artificiale vengono pagati più di un milione di dollari l'anno, si capisce che queste preoccupazioni non sono fantasia. Il problema potrebbe poi essere esacerbato dai limiti geografici e nazionali. La Cina e l'America, seguite dal Regno Unito e dal Canada, detengono il potere di sviluppo, i cervelli e le capacità per incanalare questa rivoluzione e rafforzare la loro posizione di dominio nello scenario internazionale. A noi cittadini italiani cosa resta? È necessario che passi decisi vengano compiuti rapidamente per risolvere questo divario: il rischio di essere lasciati indietro è reale e il prezzo da pagare molto alto.

Quasi l'80 per cento delle entrate di uno Stato proviene dalle tasse pagate dei lavoratori e dalle aziende. Ma se sarà un software a gestire il nostro lavoro, chi pagherà le tasse? Questa domanda sta portando studiosi, politici e capitani d'impresa a incontrarsi per dare una risposta collettiva a una questione davvero spinosa. Bill Gates ha lanciato per primo l'idea provocatoria di tassare i robot e usare quegli introiti per stimolare l'economia dell'educazione e della terza età. Ma la questione non è risolvibile così facilmente, e richiede un'analisi molto più profonda.

Più veloce e problematica sarà la disoccupazione generata dagli algoritmi intelligenti tanto più ci sarà chi ritiene che una tassa di questo tipo potrebbe essere necessaria. Ma può davvero funzionare? Per rispondere dobbiamo pensare alla questione lateralmente. Potrebbe non essere una semplice tassa, ma un minore incentivo per le aziende che sviluppano tecnologie di automazione. È quanto già succede in Corea del Sud, dove le aziende ricevono meno aiuti governativi per le tecnologie che potenzialmente compromettono posti di lavoro.

Un'altra idea è quella di dividere il guadagno incrementale generato dalle macchine tra azienda e lavoratori. Sarebbe a tutti gli effetti una tassa, ma verrebbe usata dallo Stato come sussidio di disoccupazione per i lavoratori che hanno perso il lavoro a causa della tecnologia. Ma allora dovremmo eliminare gli scavatori e tornare alle zappe? Sono tutti approcci problematici. Come definiamo cos'è un robot o una macchina in questo frangente? Chiunque usi un computer dovrebbe essere tassato? E che dire di chi installa un distributore automatico? Se lo smistamento di buste viene fatto da un robot lo tassiamo e se la busta viene dematerializzata in una email tassiamo il server di posta elettronica? Sebbene il problema possa apparire prettamente teorico, non è facile definire cosa costituisca un robot e cosa no (fatta salva la versione antropomorfizzata a cui pensiamo istintivamente).

Il trade-off della faccenda è chiaramente tra tassa e tasso d'innovazione. Il rischio infatti è che tasse più alte riducano la velocità di innovazione e questo, per un Paese, è un problema tanto grave quanto la disoccupazione. Si pensi ai Paesi emergenti nei quali le infrastrutture per le telecomunicazioni non sono sviluppate come nei Paesi occidentali.

 # 6. Le leggi in gioco con l'Intelligenza Artificiale

I veicoli autonomi evolvono, nel loro complesso, promettendo di diventare più sicuri degli agenti umani; in alcune condizioni lo sono già. Le auto autonome, essendo connesse tra loro, fanno confluire in un bacino comune tutta la loro esperienza singola a vantaggio collettivo, in modo da sfruttare i dati degli altri mezzi e accrescere così esponenzialmente il numero di scenari noti. Ciascuna automobile acquisisce così l'esperienza di guida equivalente a quella di centinaia di mezzi in viaggio per migliaia di ore per milioni di chilometri.

Dall'evoluzione dei sistemi e dall'esperienza condivisa deriva un beneficio complessivo per la società. La promessa è quella di avere meno morti sulle strade, meno danni agli autoveicoli, meno indagini e meno processi, ma anche una migliore capacità di ottimizzare situazioni critiche come le corse in pronto soccorso, simulare lo sviluppo di un'epidemia o trovarne le cure. Viene dunque da chiedersi se nel definire chi siano i soggetti che devono pagare i danni per gli incidenti, non si debba tenere in considerazione anche il beneficio collettivo che si otterrebbe da un utilizzo esteso di sistemi dotati di Intelligenza Artificiale.

Ora il punto è: possiamo individuare il colpevole dell'incidente causato da un'auto a guida autonoma basandoci sulla stessa definizione di «malfunzionamento»? Se così fosse, basterebbe estendere l'ambito di applicazione delle leggi a tutela del consumatore anche ai sistemi che utilizzano l'Intelligenza Artificiale. Tuttavia, il malfunzionamento in questione è diverso. C'è solo una predizione che porta il sistema a eseguire un comportamento difforme da quanto sarebbe auspicabile, una probabilità statisticamente minima, ma irriducibile. Tra i temi da affrontare, ce n'è uno che potrebbe rivoluzionare il rapporto tra uomo e Intelligenza Artificiale. A quali condizioni un sistema potrebbe essere dotato di soggettività giuridica e diventare un'entità i cui interessi sono tutelati dal diritto?

La soggettività giuridica è sempre attribuita alle persone fisiche (gli esseri umani), ma non è detto che in futuro non si possa parzialmente estendere anche a entità artificiali. Parallelamente, anche agli animali sono stati riconosciuti alcuni diritti, riconducibili alla loro dignità in quanto esseri senzienti dotati di emozioni, dolore, desideri e interessi, secondo una dottrina derivata dalla Dichiarazione universale dei diritti dell'animale, firmata a Parigi nel 1978. L'intelligenza (comunque diversa tra uomini e animali) non è dunque determinante per il riconoscimento dei diritti (vita e integrità fisica); semmai è pertinente per l'impossibilità di richiedere l'adempimento di doveri. Potremmo forse affermare che le persone affette da handicap cognitivi dovrebbero essere prive di diritti? Certamente no. L'intelligenza, dunque, non conferisce di per sé soggettività giuridica. Al contrario, leggendo la Dichiarazione universale dei diritti dell'uomo troviamo che «tutti gli esseri umani nascono liberi ed eguali in dignità e diritti. Essi sono dotati di ragione e di coscienza».

Seguendo questo approccio «dignitario», un sistema di Intelligenza Artificiale potrebbe ottenere soggettività giuridica ed eventualmente, capacità di agire, solo a patto che sia in grado di raggiungere un certo grado di coscienza (anche emotiva) e ragione, tale da qualificarlo come «degno» e, quindi, titolare di diritti e doveri. Si potrebbe prevedere la distruzione del mezzo e il fermo di tutti i veicoli che condividono lo stesso modello statistico. Potremmo fantasticare a lungo sull'efficacia di questa punizione rispetto alle funzioni della pena, ma un dato di fatto rimarrebbe: tanto il produttore, quanto il proprietario del mezzo resterebbero impuniti. In buona sostanza, il sogno di una soggettività giuridica per i sistemi di Intelligenza Artificiale si infrange dinanzi l'inammissibilità di situazioni in cui non si può individuare un responsabile umano: serve qualcuno che compensi i danni o subisca gli effetti di una sanzione penale.

Più appropriata al nostro contesto sembra essere la lettura secondo cui la soggettività giuridica è scissa dalla dignità. Ad esempio, il diritto riconosce capacità giuridica e d'azione a soggetti non senzienti «privi di dignità», come le società commerciali, le associazioni e le fondazioni. In questi casi, il diritto ricorre a un espediente, attribuendo soggettività a entità non biologicamente vive, ma fingendo che lo siano. Tale finzione è necessaria per consentire il raggiungimento di scopi che un solo essere umano non potrebbe mai raggiungere, come la produzione di beni in serie all'interno di un'attività d'impresa. Anche in questo caso, tuttavia, la lettura proposta non regge fino in fondo: riconoscendo soggettività giuridica all'Intelligenza Artificiale, scomparirebbe del tutto l'elemento umano, la cui presenza è assolutamente necessaria per la formazione di associazioni e imprese.

Una banca potrebbe scegliere di utilizzare trading bot spregiudicati per mettere in atto strategie d'investimento aggressive. Un giudice potrebbe definire come socialmente pericoloso qualcuno su suggerimento di un algoritmo. Un medico potrebbe fidarsi di un sistema intelligente di diagnosi dei tumori. In questi casi chi sarebbe giuridicamente responsabile se il bot perdesse il denaro, se il l'individuo socialmente pericoloso non fosse affatto tale o se il sistema di diagnosi di tumori sbagliasse? Abbiamo osservato come i sistemi di Intelligenza Artificiale prevedono forme di delegazione tali da rendere l'utilizzatore umano marginale rispetto al processo di decisione. Conosciamo, tuttavia, la necessità di garantire un controllo umano significativo. Abbiamo anche ribadito come sia poco conveniente attribuire responsabilità giuridica (liability) al sistema di Intelligenza Artificiale, sebbene questo possa essere considerato responsabile in senso causale.

Il controllo umano significativo rappresenta un fattore necessario non solo all'attribuzione di responsabilità morale (accountability), ma anche giuridica in senso stretto (liability). Sono già molti i casi in cui un soggetto è ritenuto responsabile per un fatto illecito commesso con responsabilità indiretta. Ad esempio, il genitore è responsabile per i danni arrecati dal figlio, come il datore di lavoro lo è nei confronti dei danni causati dai dipendenti. In alcuni di questi casi, come quello della responsabilità genitoriale, è anche ammessa una prova liberatoria. Si concede cioè la possibilità di dimostrare di avere adottato tutte le cautele necessarie per evitare il danno, liberando il genitore da ogni responsabilità. La scelta di un modello di responsabilità indiretta, che prevede una prova contraria, presuppone necessariamente l'esistenza di un meccanismo volto ad assicurare un controllo umano significativo, grazie al quale un essere umano possa adottare tutte le precauzioni necessarie a prevenire il danno.

Se il controllo umano significativo di un sistema di Intelligenza Artificiale non fosse presente, da un lato non si vedrebbe il motivo di concedere la possibilità di una prova liberatoria (quali precauzioni potremmo adottare?), dall'altro saremmo davanti a una piena assunzione del rischio da parte dell'utilizzatore, il quale sarebbe completamente in balia delle scelte tecniche del produttore del sistema; è necessario predisporre dei meccanismi di redress modellati sulle peculiarità del sistema coinvolto. La predizione dell'algoritmo è corretta da un punto di vista informatico: l'algoritmo esegue correttamente tutte le operazioni per cui è programmato e le sue predizioni sono da ritenere sostanzialmente affidabili, al netto dell'errore statistico implicito. Questo rende obsoleti i tradizionali metodi di verifica dello strumento tecnico.

Per valutare l'efficacia del modello alla base del quale avviene la predizione, occorre valutare la presenza di bias, la qualità dei dati di addestramento e dell'addestratore, la reazione a input non reali eccetera. Non è corretto fondare presunzioni assolute, ottenute per mezzo di sistemi di Intelligenza Artificiale. Questo è ancora più vero in assenza dei meccanismi di redress, che assolverebbero anche allo scopo di bilanciare i rischi sociali di determinate condotte, incluso l'upload di materiale online.

Diviene necessario operare su due livelli:
– creare meccanismi di governance in grado di assicurare un clima di fiducia nei confronti dello sviluppo dei sistemi di Intelligenza Artificiale (e in questa direzione va il lavoro del Gruppo Esperti ad Alto Livello sull'Intelligenza Artificiale istituito dalla Commissione Europea);
– definire in maniera democratica questi meccanismi, contemperando gli interessi degli attori coinvolti (industria, forze di polizia, cittadini, istituzioni ecc.).

Ma per vincere questa sfida è necessario dell'altro: deve essere assegnata massima priorità all'educazione ai temi tecnici, giuridici ed economici dell'Intelligenza Artificiale. Per garantire un'effettiva partecipazione democratica ai processi decisionali sul futuro dell'Intelligenza Artificiale è fondamentale spogliare questa tecnologia di quell'alone fantascientifico che mistifica la vera natura dei problemi e affrontare al contempo il quadro di riferimento giuridico ancora insufficiente e la carenza di risorse economiche destinate ad affrontare l'impatto di questa tecnologia sul mondo del lavoro.

 # Conclusioni. Alla ricerca di un equilibrio socialmente desiderabile

Un sistema di Intelligenza Artificiale perfettamente funzionante è un motore statistico, e necessariamente produce risultati probabilistici; la sua decisione potrebbe quindi essere giusta nel 98 per cento dei casi e sbagliata nel restante 2 per cento (e sarebbe inopportuno classificare le decisioni sbagliate come «errori»), il che significa che il 2 per cento delle volte una persona viene riconosciuta colpevole anche se non lo è. Per quella persona, la decisione sbagliata può avere ricadute ben oltre la portata della decisione stessa: opportunità perdute, biasimo sociale, feedback negativi online e altri effetti, che possono facilmente diffondersi nel Web e diventare impossibili da rimuovere.

Si pensi ad esempio a un'Intelligenza Artificiale che decide chi deve essere recluso e chi no. Alcune culture orientali accettano l'idea che qualche errore di valutazione, qualche innocente in prigione, giustifichi l'effetto positivo complessivo. Si potrebbe dire: «meglio un innocente in più in carcere che un colpevole in più in libertà». Nel nostro sistema di valori vale l'opposto. Nel nostro sistema democratico la valutazione del bilanciamento tra beneficio della collettività e rischio di nocumento del singolo è svolta da organismi appositi, tramite procedure rigorose.

Facciamo un esperimento mentale considerando un esempio relativo a futuribili auto a guida autonoma: l'amministratore delegato di una casa automobilistica vende i suoi prodotti (non difettosi) agli utenti, già sapendo che 10.000 di loro moriranno per una serie di motivi. La causa dei decessi non sta nei prodotti; il problema è la guida umana. L'amministratore delegato non è responsabile di questi decessi. Le persone sono responsabili. Immaginiamo ora che tale casa automobilistica introduca una futuribile tecnologia di guida autonoma che permette di abbassare il numero di vittime a 100. Sarebbe un guadagno meraviglioso per la società, ma quasi certamente le famiglie delle vittime faranno comunque causa all'azienda e denunceranno l'amministratore delegato. L'azienda potrebbe tutelarsi dal punto di vista della responsabilità civile mediante una polizza assicurativa, in modo da evitare la bancarotta a seguito dei risarcimenti per danni e responsabilità, ma l'amministratore delegato potrebbe comunque rischiare la prigione per avere immesso sul mercato un prodotto che ha causato dei decessi.

Anche per l'Intelligenza Artificiale, probabilmente, sarebbe utile un'infrastruttura normativa in grado di valutare la responsabilità delle aziende non in relazione al singolo incidente ma all'effetto complessivo, imponendo l'obbligo di fare i test appropriati, dichiarando ciò che viene ottimizzato, seguendo appropriati iter autorizzativi, legando la responsabilità alla conformità di questi protocolli istituendo organismi di vigilanza e controllo. Solo in questo modo, fissando esattamente il punto di bilanciamento tra garanzie individuali e benefici collettivi per le applicazioni che possono avere i maggiori impatti sulla vita delle persone, si possono sgravare le aziende e i loro amministratori da responsabilità che altrimenti potrebbero risultare inibenti, con l'effetto di privarci di alcuni benefici che l'Intelligenza Artificiale può apportare alla società.