

Stato dell'arte

Fin dall'inizio dello sviluppo dell'Elaborazione Elettronica fu ben chiaro che le macchine elettroniche, al pari di quelle meccaniche, avrebbero sollevato l'uomo dai compiti ripetitivi, le ultime da quelli fisici, le prime da quelle concettuali. Le macchine di entrambe le categorie ebbero un rapido sviluppo, e divennero in grado di affrontare compiti complessi, da eseguire utilizzando un *programma* scritto dall'umano, cioè un elenco di operazioni elementari organizzate in un piano sequenziale, da percorrere anche effettuando alcuni tipi di scelte, in occasione di biforcazioni del percorso esecutivo, sulla base del presentarsi di predeterminate stimolazioni.

Comunque, si trattava sempre di biforcazioni previste in anticipo, e di alternative di percorso predeterminate.

Il settore dell'elaborazione elettronica, e delle macchine in seguito denominate *robot* dovettero presto fronteggiare il seguente dilemma: “fornire loro, o non fornire, l'autonomia decisionale che gli sviluppi dell'informatica erano via via sempre più pronti a mettere a loro disposizione ?”

Il settore informatico che si occupava dell'autonomia decisionale prese il nome di AI (Intelligenza Artificiale). Dapprima analizzando, in seguito copiando, e ancora dopo potenziando le prestazioni del cervello umano, l'AI offrì un potente supporto ai processi decisionali automatici ma (per regola mai imposta, sebbene sottintesa e adottata per autoregolamentazione) *l'automatismo* venne limitato a non raggiungere il livello di *autonomia*. Significa che pur potendo i più moderni sistemi di AI apprendere dall'esperienza, arricchire le proprie capacità, suggerire comportamenti utili, eseguire compiti in contesti molto complessi e variabili, la nascente *deontologia dell'AI* **promette** di mantenere al solo umano il controllo dell'applicazione delle decisioni, al contesto reale, limitandosi ad indicare le scelte, spesso arricchendole di una chiarificazione dei livelli di efficacia.

Ma *promettere* non è sinonimo di *mantenere*. Per le motivazioni più disparate, alcune anche condivisibili, altre molto meno, già gli attuali livelli raggiunti dall'AI consentono di aggirare la decisione finale umana, e tale ramo sta vivendo velocissimi sviluppi tanto che in un futuro molto prossimo raggiungerà livelli di drammaticità tali da imporre idonee normative. A quel punto, il dramma si trasformerà in tragedia, essendo ben noto ad esempio dalla biologia quanto la regolamentazione sia intrinsecamente inadeguata, da sola, ad affrontare i crescenti problemi di bioetica.

Al momento attuale, solo una crescente consapevolezza, da parte degli utilizzatori, di potenzialità e rischi dei sistemi di AI potrà contribuire a delimitarne i confini mediante esercizio del potere di opinione e, soprattutto, punendo sul mercato le applicazioni sgradite. Purtroppo, esistono ampi ambiti che non sono sotto il controllo

dell'utente (ad es. governativo, militare, spaziale, della ricerca, finanziario ecc.) e il problema permane quasi integro.

Per meglio commentare la situazione già attuale e sicuramente futura, è opportuno presentare brevemente potenzialità e rischi delle imminenti applicazioni, corredandole di una analisi sociologica, riportata sulla bibliografia allegata, di sondaggi in merito ai timori dell'utenza. Solo per limitarsi ad alcune categorie rilevanti, si va dalla guida di veicoli assistita da AI, alle operazioni chirurgiche, alle diagnosi mediche, alla cura delle lesioni cerebrali, alla didattica, all'analisi indiziaria nel campo della giustizia, al giudizio processuale, al controllo di veicoli spaziali o di armamenti militari, a robot per mansioni assistenziali, infermieristiche, di sorveglianza, alla profilazione di persone pericolose, alla finanza aggressiva, alla propositività nel marketing, alla violazione della privacy e nella sicurezza, al furto d'identità, ai sistemi per il miglioramento delle prestazioni su determinati compiti, fino alla manipolazione delle masse e ai sistemi di persuasione in ambito commerciale o comportamentale.

Una regolamentazione è solo un ovvio passo preliminare necessario, ma assolutamente insufficiente, data la dirompente capacità di tali sistemi, mai precedentemente sperimentata nell'esperienza umana. Gli attuali sistemi dispongono già della capacità (per ora a potenziale) di scatenare (perfino volutamente) una guerra tra stati di livello tale da mettere a rischio la sopravvivenza dello stesso genere umano o di asservirlo ad un *ordine artificialmente deciso*, **sostituendolo!** Un tale scenario potrebbe essere innescato anche, inizialmente, dall'intelligenza umana deviata, ma questo aspetto non diminuirebbe la gravità del conseguente danno. Come avviene per i virus biologici costruiti in laboratorio e, più o meno accidentalmente, sfuggiti al controllo, virus informatici possono già oggi essere costruiti in modo da violare i sistemi informatici, riprodursi, prendere via via il controllo a scala maggiore. Inoltre alcuni esempi di applicazioni dell'AI già oggi disponibili illustrano senza ombra di smentibilità di essere progettati con la *capacità* (forse non con l'intenzione) di operare senza il controllo umano.

Una prima classe di esempi contiene intrinsecamente, nel proprio obiettivo dichiarato, tale aggiramento. Si tratta delle *protesi cerebrali*, da impiantare nel cervello di pazienti che hanno riportato lesioni in alcune regioni. Esse raccolgono, mediante elettrodi, segnali provenienti dalle regioni collegate, li elaborano con l'AI e restituiscono segnali *imperativi* ai neuroni cerebrali, ripristinando le funzioni dell'individuo. Si tratta di applicazioni invasive (vedere brevetti _ _ _ _ _) circa le quali non si sollevano obiezioni etiche, ma che pur tuttavia mostrano come il controllo cerebrale venga aggirato di fatto. Già si ventilano impianti, destinati a soldati del futuro, che ne aggirino le facoltà decisionali, potenziandole, ma di fatto sottraendone il controllo all'individuo, delocalizzando il processo decisionale e demandandolo ad umani remoti che telepilotano il soldato, se non a sistemi di sola AI. Talvolta tali sistemi presentano un livello di invasività tale da richiedere che una

vasta parte della calotta cranica venga sostituita chirurgicamente da una calotta artificiale da impiantare sotto la cute del capo, con elettrodi inseriti direttamente nell'encefalo, in comunicazione con apparati di AI in parte alloggiati nell'aumentato spessore della calotta, in parte esternamente, raggiungibili in collegamento wireless (vedasi brevetto _ _ _ _). Un ben noto progetto di ricerca, che promette rapidi sviluppi, denominato Neuralink (_ _ _ _) offre un'interfaccia invasiva tra cervello umano e sistema esterno di AI, dotata della capacità di stimolare direttamente neuroni del cervello biologico (_ _ _ _ _). L'opinione pubblica attuale accetta con più favore i sistemi atti al ripristino di funzionalità perdute, e mostra notevole allarme nei confronti di quelli per potenziare funzioni già esistenti, chiedendosi quali garanzie essi offrano di non sopraffare l'individuo. Per quanto riguarda la guida automatica di veicoli, più diffusa nella realtà di quanto non venga pubblicizzata, vanno distinte due categorie: i veicoli a guida automatica per la movimentazione di pezzi nelle catene di montaggio (senza passeggeri a bordo) e quelli destinati a controllare automobili e treni con persone a bordo. Nel primo caso gli incidenti agli umani sono limitati al personale eventualmente investito, mentre nel secondo riguarderebbero il personale trasportato. L'opinione pubblica accetta i primi, ma ha forti timori a riguardo dei secondi.

Purtuttavia gli studi stanno procedendo velocemente e spesso l'opinione pubblica non viene sempre esplicitamente informata: quando mai i piloti d'aereo informano i passeggeri dell'inserzione del pilota automatico, quandanche con la supervisione umana? E tale supervisione resta sempre sufficientemente attenta? I veicoli con posteggio assistito o automatico sono un chiaro esempio di come l'autonomia dell'AI si faccia strada in modo strisciante, non eclatante, insinuandosi nelle abitudini, aggirando la fase di vaglio razionale prima di una accettazione consapevole. Il potenziamento della *mobilità* della mano del chirurgo mediante robot dedicati alla esecuzione di interventi è senz'altro un potente ausilio nel settore sanitario, al momento completamente sotto controllo umano, una potenzialmente suscettibile di maggiore autonomia da parte della macchina, e ciò deve venire crescentemente attenzionato. La diagnostica medica assistita da AI, come parallelamente le inferenze indiziarie che precedono un processo, lasciano attualmente il giudizio finale all'umano: medico o giudice rispettivamente, ovvio; ma meno ovvio è chiedersi fino a che punto le informazioni fornite dall'AI acquistino agli occhi del giudicante finale un'autorevolezza tale da viziare la libertà di giudizio... A riguardo, l'opinione pubblica accetta con più favore le applicazioni nel settore sanitario rispetto a quelle nel settore della giustizia, ad es. a riguardo delle tecniche di riconoscimento facciale, degli screening di massa, dell'identificazione mediante parametri biometrici, di profilazione generalizzata ecc. Il campo didattico è già costellato di tutorial per insegnare determinate competenze e sistemi di valutazione automatica dei livelli conseguiti. L'AI è certamente un utile strumento, ma rischia di mettere in secondo piano l'apprendimento *formativo* contraddistinto dal confronto tra pari, dalla comparazione per analogie, dalla discussione delle ipotesi e della valutazione consapevole dei percorsi logici. Persino l'utilissimo appoggio fornito dall'AI in termini di ricerca di informazioni per argomento, vincente in termini di velocità ed

esaustività, pecca dai lati della pertinenza e della affidabilità delle informazioni reperite, rispetto al personale cammino di indagine, tra l'altro pieno di potenzialità di arricchimento formativo durante il percorso stesso. Demandare quindi, eventualmente, in futuro la maggior parte della didattica a sistemi di AI sarebbe un grave sbaglio che si pagherebbe in termini di discenti *ammaestrati* invece che *formati*, poco in grado di incidere sul progresso umano. Nel controllo di veicoli spaziali, attualmente alcune decisioni (da prendere immediatamente) vengono lasciate all'autonomia del sistema, mentre quelle per cui è consentito un tempo più ampio (solitamente anche di maggiore profilo strategico) possono venire suggerite dall'AI al controllo umano, ma è quest'ultimo che ha la parola finale.

Paragonando il settore spaziale a quello militare, si nota che al fallimento di una missione spaziale si attribuisce molto più valore rispetto, ad esempio, al danno collaterale di distruzione di un ospedale... La comparazione delle logiche dei due settori è piuttosto complessa: mentre si dà grande valore al sistema spaziale (astronave, astronauti, tempo investito ecc.) il valore, ad es. di un drone è limitato sia nel costo, nel personale a bordo (assente) e nel tempo di approntamento. Si è perciò disposti a lasciare all'AI le decisioni tattiche, perché attimi di indecisione farebbero fallire la missione. Invece la strategia è sotto controllo umano, essenzialmente perché non viene, di fatto, considerata di prioritario rilievo evitare danni collaterali al nemico.

Nel caso di robot utilizzati per dare informazioni nei pubblici uffici, o per assistenza agli anziani o per compiti di sorveglianza, le problematiche etiche sono sostanzialmente diverse. Di solito non viene percepito come illecito (solo molto fastidioso) che l'obiettivo principale del fornitore di informazioni al pubblico sia, di fatto, esaurire velocemente la coda smistando il richiedente verso altra sorgente di informazione (ad es. facendogli perdere tempo a registrarsi, accedendo ad un portale, inoltrando una richiesta online ecc. Non necessariamente è illecito, comunque tale comportamento sposta l'impegno del tempo dall'ente al fruitore, causandogli (a conti fatti) una perdita valutabile anche in termini economici. Molto spesso le azioni che l'utente viene deviato a fare lo costringono ad un inutile loop dal quale esse solo abbandonando il tentativo di risolvere il problema. Se si tratti o meno di un *bug* del sistema è comunque imperdonabile. Già come è ingiusto, nascondendosi dietro l'obiettivo di personalizzare le informazioni da fornire, raccogliere sull'utente dati sovrabbondanti, profilandolo. L'opinione pubblica però, è molto più critica nei riguardi dell'assistenza infermieristica da parte di AI, soprattutto per gli aspetti di somministrazioni di terapie. Va fatta qui una chiarificazione: appare molto evidente che l'umano, pur riconoscendo ai sistemi di AI una minore frequenza di errore rispetto ad operatori umani, *preferisce* comunque che a sbagliare sia eventualmente un altro umano: tale comportamento sembra sottintendere di non essere interessato alla pura statistica, ma a quelle doti di empatia col proprio simile che mancano nel rapporto con la macchina. Proprio per questo motivo i più recenti sistemi per l'assistenza si stanno sviluppando nella direzione di condividere con l'assistenza

anche situazioni di coinvolgimento emotivo (_____).

Nel caso della sorveglianza, un alto livello di autonomia del sistema di AI può condurre ad azioni violente nei confronti del trasgressore difficilmente accettabile per legge, quando eseguite da un sistema *freddo*: si pensi al criterio di legittima difesa secondo il quale l'agredito sente in pericolo la propria vita, evidentemente inapplicabile ad un sistema artificiale. Per questo i sistemi ad alta autonomia sono molto sgraditi all'opinione pubblica.

Nella profilazione per prevenire crimini, l'etica impone dei limiti che dovrebbero solo consentire di raccogliere elementi atti alla prevenzione dei reati, per sventarne l'esecuzione, ma mai esercitare violenza punitiva esclusivamente su base indiziaria. La finanza aggressiva, capace di spingere i mercati, mediante studi di previsione effettuati con sistemi di AI, verso direzioni innaturali se valutate con le dinamiche umane, provoca *cadaveri trasparenti*: fallimenti, perdite di posti di lavoro, suicidi, sovvertimento di equilibri economici, fino a volatilizzazione di capitali e tensioni internazionali. Sistemi per il *marketing persuasivo* spesso non si limitano a proporre la propria offerta secondo una profilazione delle presumibili necessità dell'utenza, ma si pongono in sintonia emotiva con questa, sollecitandone i lati vulnerabili, ventilando addirittura occulte *punizioni morali* a carico della coscienza del soggetto attenzionato, lasciando intuire una sorta di rilassamento della tensione emotiva in caso di perfezionamento dell'acquisto. Non si ritiene qui necessario sottolineare quanto questa crescente pratica richieda un urgente antidoto. Le continue violazioni della privacy personale (fino al furto d'identità) e della sicurezza informatica dei sistemi organizzati della società, a cui continuamente assistiamo, pur non essendo per ora riconducibili all'autonomia comportamentale di sistemi di AI impazziti, ma ad attività di hacker individuali o organizzati, persino per fini politici e bellici, indicano tuttavia le potenzialità di danno che, in mano a sistemi autonomi di AI potrebbero essere messe in atto rapidamente, potentemente e inesorabilmente.

Non tutte le applicazioni dell'AI sono però criticabili, piuttosto la critica è legata all'uso e ai confini applicativi. Il miglioramento delle prestazioni umane nell'espletamento di determinati compiti beneficia, ad esempio, notevolmente del supporto dell'AI in praticamente tutti i settori. Va ribadito che va prestata comunque attenzione al fatto che si tratti di *formazione* e non di *ammaestramento*, per lasciare il soggetto didatta nell'ambito della dimensione umana. Da ultimo, ma forse è l'oggetto più importante, la profilazione dinamica di massa è in grado di indicare ai sistemi di AI non soltanto quali siano i bisogni (non solo commerciali) percepiti rilevanti dall'utenza, ma i ben più importanti legami di causa-effetto tra la disponibilità di nuove risorse (culturali, materiali, di benessere fisico ed economico) offerte o offribili dal mercato e stimarne l'impatto in termini di trasformazione di abitudini, priorità, perfino valori morali. Tutto ciò costituisce le *manopole* che possono consentire a sistemi complessi di AI di *regolare* i grandi cambiamenti della società umana sottraendone il corso e le decisioni al genere umano stesso, manipolando in modo

invisibile il comportamento delle masse. Già oggi lo si sta facendo, sulla base di decisioni prese da un manipolo di grandi attori dello scenario finanziario e commerciale, che reggono i fili dei *burattini* capi di stato. Presto, a tali disumani-umani, potrebbero sostituirsi AI didumane-disumane.

Per concludere questa panoramica, che ha lo scopo di individuare le criticità contenute nei tanti problemi rappresentati, si descrive una invenzione che impernia la propria **UTILITA'** nel riproporsi di soddisfare le principali necessità comuni all'articolato quadro esposto:

- potenziare l'intelligenza umana, in confronto con le potenzialità delle macchine esterne;
- schermarla da sgraditi accessi provenienti dall'esterno;
- far filtrare i soli ausili ritenuti vantaggiosi.

La **NOVITA'** dell'invenzione risiede nell'assoluta mancanza, ad oggi, di difese verso il citato quadro.

La **REALIZZABILITA'** dell'invenzione viene soddisfatta mediante la descrizione delle **idee** che la supportano; ciò consente ad esperti del settore di assemblarla e programmarne l'operatività, nonostante essa non sia approcciabile con criteri di sola progettazione, in mancanza delle citate idee.

Prof. Ugo Abundo:

fondazioneneumann.org