

Il fine dell'Intelligenza Artificiale

Francesco Follo

Il tema dell'Intelligenza Artificiale cattura oggi l'impegno e la riflessione di molti uomini di scienza a livello internazionale, che afferiscono a diverse discipline: ingegneria, logica e matematica, neuroscienze e la filosofia. È anche molto presente nell'immaginario collettivo, come testimoniato persino dalla vasta filmografia in merito. Questo evento promosso dall'UNESCO è un chiaro segno dell'importanza del tema in questo momento storico.

Potrà una macchina pensare – nel senso pieno del termine? Potrò in tutto e per tutto equipararsi ad un essere umano con la sua vita intellettuale integrale? Questi sono evidentemente alcuni tra gli interrogativi più profondi e radicali che gli sviluppi dell'Intelligenza Artificiale pongono all'umanità. Essi riguardano non solo gli uomini di scienza, ma anche politici e sociologi, filosofi e uomini di fede.

La storia dell'Intelligenza Artificiale – che in senso proprio inizia con gli anni '50 del secolo scorso – vede alternarsi fasi di grande ottimismo e fasi di presa di coscienza delle difficoltà nel riprodurre l'intelligenza "naturale", quella umana. Non è questa la sede per ripercorrere tale storia. È significativo, però, sottolineare che la valutazione dei risultati raggiunti dipende fortemente dagli obiettivi che la comunità scientifica e tecnologica si pone di volta in volta. Da questo punto di vista, rimane importante la distinzione tra "Intelligenza Artificiale forte" e "Intelligenza Artificiale debole". L'Intelligenza Artificiale debole si pone come fine la progettazione e costruzione di macchine che agiscano *come se* fossero intelligenti. L'Intelligenza Artificiale forte, invece, si pone l'obiettivo massimale di concepire macchine effettivamente simili all'uomo, al punto da poter sviluppare persino un'autocoscienza. Inoltre, secondo alcuni ottenere un'Intelligenza Artificiale implicherebbe riprodurre in tutto e per tutto il substrato materiale che assicura all'uomo la sua pratica razionale: il cervello. Secondo altri, invece, si tratterebbe "semplicemente" di riprodurre l'esito (o alcuni degli esiti) di quella pratica razionale, indipendentemente dalle specifiche fisiche e ingegneristiche escogitate per la struttura della "macchina pensante".

A tutt'oggi, una macchina che riproduca i dettagli dell'organizzazione cerebrale e che sappia esibire tutte le caratteristiche più alte dell'intelligenza umana, autocoscienza inclusa, sembra al di là della portata degli sviluppi tecnici correnti o concretamente prevedibili. A tale proposito, va anche tenuto presente che, malgrado i grandissimi successi delle neuroscienze, un'immagine chiara e completa di come sia costituito e di come funzioni il cervello non è ancora disponibile. Tuttavia, macchine (o programmi informatici) in grado di compiere operazioni complesse con prestazioni comparabili – o talvolta, per specifici contesti, persino superiori – a quelle umane – esistono e, in alcuni casi hanno anche lasciato il laboratorio di ricerca per diffondersi in molteplici applicazioni tecniche sia di frontiera, sia ormai commercialmente disponibili. Questo anima un ottimismo di fondo che lascia aperta, almeno per alcuni, la possibilità – non importa quanto remota – di arrivare un giorno ad esaudire i desideri dell'Intelligenza Artificiale forte.

Le brevi note che seguono sono volte a riflettere su alcuni aspetti che – a mio parere – potrebbero, in qualche modo, sancire differenze qualitative o "essenziali" tra Intelligenza Artificiale e intelligenza umana. Si tratta di riflessioni filosofiche informate su alcuni degli aspetti tecnico-scientifici coinvolti.

Il problema appare certamente meno profondo se si considera l'Intelligenza Artificiale debole, dal momento che, per definizione, in questo caso l'obiettivo è quello di produrre macchine in grado di operare *come se* fossero intelligenti, senza pretendere di essere vere e proprie riproduzioni dell'intelligenza umana. Inoltre, i risultati già ottenuti – talvolta realmente sorprendenti – si rivolgono comunque a contesti e compiti specifici e limitati. Tuttavia, se gli sviluppi futuri fossero effettivamente in grado di produrre macchine che operassero come se fossero intelligenti in *ogni contesto* o situazione, e a proposito di *qualsiasi ambito* di pensiero e azione tipicamente umani, ciò ci confronterebbe di per sé con uno dei due aspetti dell'Intelligenza Artificiale forte – ovvero la possibilità di riprodurre *in toto*, l'intelligenza umana. È vero: in un simile caso, non si avrebbe ancora la realizzazione del sogno di riprodurre

non solo le prestazioni, ma anche la “struttura” o la “costituzione” del substrato dell’intelligenza umana. Tuttavia, considerando che, come accennato, oggi non possediamo ancora un’immagine chiara e completa di come sia costituito e, soprattutto, di come funzioni il cervello, la realizzazione di macchine in grado di operare *come se* fossero intelligenti in *ogni contesto e qualsiasi ambito*, porrebbe comunque in maniera radicale il problema dell’unicità e della irriducibilità dell’intelligenza umana.

Vi sono, secondo me, almeno tre aspetti che rendono difficile pensare che un giorno le macchine potranno sostituire in tutto e per tutto l’intelligenza umana: la dimensione affettiva, la dimensione semantica, e una terza dimensione che, per ora, chiamerei “motivazionale”.

La tradizione filosofica ha sempre riconosciuto, in un modo o nell’altro, l’influenza di affetti, emozioni, sentimenti, e “appetiti” sulla cognizione umana. Oggi sembra chiaro anche dal punto di vista sperimentale che gli stati emotivi influiscono sui processi decisionali e razionali: sulle scelte. Questo, peraltro, anche indipendentemente dal tema dell’Intelligenza Artificiale, pone un problema dal punto di vista della cosiddetta “rational choice theory”. Secondo questo approccio, una decisione razionale scaturirebbe da un processo interamente logico che tende ad una decisione ottimale – vale a dire, che massimizzi l’utilità e minimizzi i rischi. I processi decisionali umani reali – anche, ma non soltanto, per la rilevanza degli stati emotivi per come sottolineato dalla cosiddetta “affect heuristics” – non seguono sempre e necessariamente un tale processo. Neppure gli obiettivi delle persone in carne ed ossa sembrano sempre essere quelli della massimizzazione dei risultati, come sottolineato dal fatto che i decisori reali possono anche accontentarsi di soluzioni “sufficientemente buone” pur se non ottime (si consideri la distinzione tra “massimizers” e “satisficers”). Se le cose stanno così, allora un’Intelligenza Artificiale (forte) che intendesse riprodurre in tutto e per tutto l’intelligenza umana, dovrebbe riprodurre anche gli aspetti emotivi e affettivi. Ora, in primo luogo, non sembra oggi possibile riprodurre artificialmente quella sfera emotiva e “sentimentale” che caratterizza diffusamente l’intelligenza umana: e questo pone già un limite importante al sogno dell’intelligenza Artificiale (forte). In secondo luogo, introdurre questa dimensione nelle “macchine pensanti” renderebbe il loro procedere meno “razionale” (nel senso della “rational choice theory”). Queste considerazioni pongono un interrogativo profondo a coloro che, dal punto di vista tecnico, si occupano di Intelligenza Artificiale: quale deve essere il fine di questi sviluppi tecnologici? Riprodurre e imitare l’intelligenza umana, oppure supportare quest’ultima in contesti ed ambiti *specifici* lasciando gli aspetti integrali e complessivi all’uomo in carne ed ossa, ragione e sentimento?

A tale proposito, poi, vorrei notare che la tradizione filosofica (e anche teologica) ha sempre sottolineato l’importanza di amore e carità per la ricerca stessa della verità. Perseguire fini conoscitivi alti, richiede una tensione verso la verità che non può essere interamente spiegata in termini utilitaristici o “di convenienza”. Questo è un aspetto di quella dimensione motivazionale prima accennata e sulla quale tornerò a breve. Per ora, vale la pena sottolineare che una macchina intelligente che volesse imitare e riprodurre esaustivamente l’intelligenza umana dovrebbe contemplare anche questo aspetto direi “emotivo alto”, accanto naturalmente a quanto già detto rispetto alla generale influenza degli stati emotivi sui processi cognitivi e decisionali.

Un secondo aspetto problematico dell’Intelligenza Artificiale (forte) ha a che fare con la distinzione – ormai classica nella filosofia del linguaggio moderna e contemporanea – tra sintassi e semantica. È opinione diffusa che i calcolatori, compresi quelli che dovrebbero supportare una intelligenza artificiale, possano trattare soltanto legami sintattici tra simboli non interpretati (cioè privi di significato), mentre non potrebbero trattare i contenuti semantici possibilmente attribuibili a tali simboli. Questo, di per sé, porrebbe un enorme limite alla possibilità di realizzare un’Intelligenza Artificiale che possa imitare *in toto* l’intelligenza umana, la quale trova nell’universo dei significati una dimensione fondamentale da molti punti di vista. A tale proposito viene talvolta proposta l’idea che, in realtà, anche la semantica di alto livello potrebbe essere ridotta alla sintassi, che il limite sarebbe solo quello della potenza di calcolo e dell’enorme complessità necessaria, e che quindi in futuro si potrà giungere a macchine in grado di trattare anche gli aspetti semantici più sofisticati esibiti dall’intelligenza umana. A tale proposito, credo che due considerazioni potrebbero risultare opportune e interessanti.

In primo luogo, la linguistica ha spesso considerato la sintassi e la semantica come due aspetti distinti del linguaggio. La sintassi definirebbe una serie di regole di composizione generali applicabili a ampie classi di elementi linguistici senza riguardo al significato che risulterebbe da quelle costruzioni composte. Approcci più recenti – sempre più

connessi con gli sviluppi delle neuroscienze –propongono invece una maggiore interdipendenza tra sintassi e semantica (per esempio, le cosiddette “Cognitive Grammar” e “Construction Grammar”). Il punto rilevante è che questi sviluppi suggerirebbero una “riduzione” della sintassi alla semantica piuttosto che viceversa! In altre parole, le costruzioni sintattiche usate nel linguaggio e nel ragionamento umani sarebbero intrinsecamente dipendenti dai concetti e/o dai significati delle espressioni combinate – vale a dire, dalla semantica. Di conseguenza, secondo questi approcci, non esisterebbe una “sintassi separata” genericamente applicabile agli elementi semantici indipendentemente dal loro significato.

In secondo luogo, non andrebbe perso di vista cosa sia la semantica. Spesso si pensa che la semantica sia soltanto la rete di relazioni tra termini. Ad esempio, se si cerca su un dizionario un termine, questo termine è definito da altri termini messi in relazione. “Calendario” è definito come un insieme di fogli che riportano i giorni, le settimane e i mesi in un anno. Chi conoscesse il significato di tutti i termini usati in questa definizione potrebbe capire il termine “calendario”; chi non ne conoscesse alcuni, potrebbe continuare a cercare sul dizionario, ricorsivamente. Tuttavia, per quanto si portasse avanti questo processo, l’individuo che non avesse mai visto o posseduto un calendario, assai difficilmente avrebbe una autentica comprensione di quel termine. La comprensione di un termine dipende profondamente anche dalle *esperienze reali* fatte dal parlante nella sua vita. Il significato di parole come “povertà” o “libertà” cambia molto a secondo della situazione personale, della propria storia di vita e anche del contesto storico e/o geografico generale in cui ci si trova. I significati sono connotati emotivamente tanto quanto lo sono razionalmente. Su questa base, dunque, realizzare macchine che possano – secondo l’obiettivo dell’Intelligenza Artificiale forte – riprodurre ogni aspetto dell’intelligenza umana non avrebbero solo bisogno di una sufficiente potenza di calcolo ma, per dirla enfaticamente, avrebbero bisogno di *vivere* come un essere umano. Non solo computare simboli, ma fare esperienze: patire e gioire, desiderare e temere, vedere, udire, toccare, odorare e gustare. A questo punto, dunque, l’Intelligenza Artificiale implicherebbe anche una “Vita Artificiale” e, come sapete, i problemi incontrati nei tentativi di produrre “macchine viventi” sono almeno altrettanti, e altrettanto gravi, di quelli incontrati dall’Intelligenza Artificiale.

Giungerei così al terzo aspetto che mi sono riproposto di toccare, quello che prima ho chiamato “motivazionale”. È chiaro, oggi, che molti aspetti dell’intelligenza umana provengono dal lungo percorso evolutivo che ha portato alla nostra specie biologica. Alcuni di questi aspetti sono anche condivisi – almeno in parte – da altre specie animali non umane. È anche noto che, secondo le teorizzazioni circa l’evoluzione biologica, le novità che sono emerse nella storia naturale, anche dal punto di vista comportamentale e cognitivo, rispondono ad una logica di necessità. Necessità intesa non nel senso che il processo evolutivo sarebbe in qualche modo necessitato, ma nel senso che lungo la storia evolutiva emerge ciò che, per le varie specie biologiche nei vari contesti ambientali, è necessario per sopravvivere – pena, appunto, soccombere alle sfide ambientali. Questo quadro concettuale, sebbene certamente valido per molti aspetti persino della cognizione umana, sembra però non essere in grado di spiegare interamente quella che a buon diritto può essere chiamata l’evoluzione culturale caratteristica della storia umana. La domesticazione di piante e animali, la costruzione di città e luoghi di culto, l’invenzione della scrittura e dell’aritmetica, la nascita della cosiddetta cultura teorica, delle università, della scienza moderna, le rivoluzioni industriali che hanno segnato gli ultimi secoli: tutto ciò è difficilmente ascrivibile solamente alla necessità nel senso di cui poco sopra. Nessuna di queste innovazioni (e le miriadi di invenzioni specifiche che le hanno accompagnate), strettamente parlando erano necessarie alla sopravvivenza dell’essere umano. La questione risulta ancor più profonda se guardata dal punto di vista dei singoli individui che hanno dato contributi essenziali a questi avanzamenti. Si pensi a Socrate, che per amore della verità e della giustizia si è lasciato costringere al suicidio; oppure a Galileo Galilei che nella convinzione delle sue idee cosmologiche è andato incontro a due processi. Queste poche e sommarie riflessioni – sulle quali molto potrebbe essere aggiunto – pongono con forza il problema delle *motivazioni* che spingono l’essere umano a ricercare, a voler conoscere ed inventare – che lo spingono verso la verità e a voler migliorare le proprie condizioni materiali e spirituali al di là delle strette necessità.

Questo è un tema grande, che certamente non può essere esaurito in questa sede. Tuttavia, esso pone un ulteriore problema all’obbiettivo dell’Intelligenza Artificiale forte. Una macchina pensante che volesse riprodurre *in toto* l’intelligenza umana dovrebbe anche saper riprodurre questo aspetto “motivazionale”. In altre parole, non dovrebbe solamente essere in grado di compiere operazioni intelligenti per risolvere compiti etero-assegnati, ma dovrebbe

essere anche in grado di auto-assegnarsi compiti e traguardi, di avere aspirazioni. Oggi, reti-neurali sofisticate e i cosiddetti “sistemi esperti” sono in grado di compiere operazioni intelligenti in maniera comparabile – e talvolta persino superiore – rispetto all’essere umano. In alcuni casi, possono farlo persino escogitando soluzioni non previamente immesse nel sistema: si pensi ai cosiddetti “algoritmi genetici”, vale a dire, programmi capaci di modificarsi autonomamente in modo da migliorare le proprie prestazioni. Tuttavia, non soltanto possono fare questo solamente in domini limitati e circoscritti ma, soprattutto, non si assegnano autonomamente gli obiettivi da raggiungere. Una reale Intelligenza Artificiale forte dovrebbe, invece, essere in grado di far questo. Non a caso, molti film sull’argomento pongono, in un modo o nell’altro, il tema della macchina che si ribella, che vuole essa stessa auto-determinarsi, o che desidera “diventare umana”, o che intende proteggere l’umanità piuttosto che soggiogarla e conquistare la Terra. Non mi pare che alcuno degli sviluppi tuttora in corso o realisticamente prevedibili possano realmente portare ad un simile esito.

In conclusione, vorrei offrire ancora altre due brevi riflessioni conclusive. La prima è che evidentemente i tre aspetti che ho voluto affrontare sono strettamente connessi tra loro e, benché io ritenga che l’ultimo taccato sia il più profondo, sembra chiaro che le motivazioni sono strettamente connesse sia con la dimensione affettiva sia con quella semantica. La seconda riflessione conclusiva intende accennare ad un tema filosofico-teologico centrale nella tradizione cristiana: la questione dell’anima umana. Non intendo qui affrontare la questione in tutto il suo respiro, ma soltanto notare che il tema della motivazione che spinge l’uomo ad auto-determinarsi e ad auto-superarsi continuamente, anche al di là delle necessità, ha dei legami assai stretti con alcune delle “funzioni” che la tradizione filosofica e teologica cristiana attribuisce all’anima: quelle riguardanti la libertà, la dignità, l’autocoscienza e la coscienza morale. È quindi degno di nota che alcuni dei problemi che sgorgano dagli obiettivi – e degli interrogativi – posti dall’Intelligenza Artificiale possano trovare dei corrispettivi, delle risonanze, e delle ragioni di approfondimento anche in alcune tematiche filosofico-teologiche tradizionali di stampo eminentemente antropologico.