

Intelligenza e comprensione: confronto fra naturale e artificiale

Luca Mari

L'intelligenza, anche quella non-artificiale, è un'entità complessa. Ne proponiamo qui un'esplorazione, alla luce delle domande se, quando, e a quali condizioni un sistema artificiale possa esibire un comportamento intelligente, e orientata a comprendere gli scenari che potrebbe prospettare la sempre più ampia diffusione di sistemi artificiali che manifestano comportamenti sempre più intelligenti.

Un agente artificiale può esibire un comportamento intelligente, e nel caso quando e a quali condizioni? Una breve analisi e qualche riflessione sulla sempre più ampia diffusione di sistemi artificiali che manifestano comportamenti sempre più intelligenti.

Intelligenza: un'entità complessa

Intelligenza: già l'ambiguità dell'etimo coglie la complessità del concetto: *intus-legere*, "leggere dentro", dunque capacità analitica e di approfondimento, e *inter-legere*, "leggere attraverso", dunque capacità associativa e di connessione. Questa ambiguità diventa molteplicità, prima di tutto nel senso categoriale di Howard Gardner, secondo cui esistono diverse e relativamente indipendenti tipologie con cui l'intelligenza si manifesta¹, ma anche in senso operativo come capacità di risolvere problemi, dove la molteplicità in tal caso sta nel fatto che esistono tipi di problemi diversi: vincere nel gioco degli scacchi è una capacità diversa da guidare in modo appropriato un'automobile. La capacità focalizzata alla soluzione di una sola classe di problemi è già intelligenza? L'intelligenza dipende dal contesto? Si può essere intelligenti senza una prospettiva sufficientemente ampia? Senza avere un qualche tipo di autonomia decisionale? (qualunque cosa ciò significhi) Senza avere intelligenza anche di se stessi? (significativamente, una tipologia di intelligenza secondo Gardner è l'intelligenza intrapersonale, dunque la capacità di conoscere se stessi).

Senza un modello di sfondo, che dovrebbe essere in grado di fornire una caratterizzazione della *natura* dell'intelligenza, non è chiaro come giustificare una risposta a queste domande – che anche qui rimarranno sullo sfondo – e si tende a un esito di tipo operazionistico: si ammette che 'intelligenza' è un concetto aggregato (un *cluster concept*), e ogni volta che usiamo il termine "intelligenza" lo usiamo come sinonimo di "intelligenza-di-tipo-X", per uno o più X, il cui significato è 'capacità di risolvere problemi di tipo X' (il caso canonico è X=test per quoziente di intelligenza).² Di questo territorio proponiamo qui una breve esplorazione,

¹ Una breve introduzione al tema, proposto nel libro di H. Garden, *Formae mentis – Saggio sulla pluralità dell'intelligenza*, Feltrinelli, Milano, 1987, è in it.wikipedia.org/wiki/Intelligenza#La_teoria_delle_intelligenze_multiple.

² L'operazionismo (sviluppato a partire dal libro di P.W. Bridgman, *La logica della fisica moderna*, Bollati Boringhieri, Torino, 1952 – prima edizione originale 1927; si veda per esempio it.wikipedia.org/wiki/Operazionismo) è un grimaldello che si propone di aprire tutte le porte del genere

orientata a comprendere gli scenari che potrebbe prospettare la sempre più ampia diffusione di sistemi artificiali che manifestano comportamenti sempre più intelligenti.

Da intelligenza naturale a intelligenza artificiale

Il rimedio all'assenza di un modello di sfondo convincente, e un tentativo di alternativa all'operazionismo, è stata per lungo tempo l'ipotesi che l'intelligenza, non diversamente dalla capacità di gestire informazione, sia più o meno per definizione una proprietà degli esseri umani, eventualmente estesa ad altre entità biologiche, "superiori" appunto perché dotate di qualche forma o qualche grado di intelligenza. Secondo l'Enciclopedia Britannica, per esempio, l'intelligenza artificiale è "l'abilità di un calcolatore digitale o di un robot controllato da un calcolatore di eseguire compiti comunemente associati a esseri intelligenti".³ In questo senso, il concetto di intelligenza artificiale sarebbe metaforico: un agente⁴ artificiale che si comportasse in modo indistinguibile da un agente naturale considerato intelligente in questo comportamento non sarebbe comunque "davvero" intelligente, perché con il suo comportamento solo *simulerebbe* intelligenza. Ma, in assenza di un modello di sfondo a cui ancorarsi, questa posizione diventa rapidamente poco convincente: proprio come accade per la capacità di gestire informazione, si può riconoscere che l'intelligenza non è qualcosa che ha senso simulare. Simulare di fare un calcolo o di suonare una sinfonia non è diverso da fare quel calcolo e suonare quella sinfonia. E non si capisce in che senso si potrebbe sostenere che un programma che gioca a scacchi non gioca davvero, ma solo simula di giocare. Per questo motivo, l'intelligenza simulata è a tutti gli effetti indistinguibile dall'intelligenza. La conseguenza, non banale, è che, in assenza di un modello di sfondo l'intelligenza è una proprietà non che si ha, ma *di come ci si comporta*.

E' questa la giustificazione del test⁵ che Alan Turing propose per caratterizzare un comportamento intelligente: data l'ipotesi che l'intelligenza sia una proprietà che gli esseri umani generalmente esibiscono, qualsiasi comportamento indistinguibile da quello, riconosciuto appunto come intelligente, di un essere umano è a sua volta intelligente. E del resto ciò è consistente con l'uso che facciamo abitualmente del concetto: consideriamo intelligente una persona che si comporta generalmente in modo intelligente, ma anche una persona che consideriamo intelligente può a volte comportarsi in modo stupido, e se un comportamento è stupido lo è anche se esibito da una persona abitualmente intelligente. Dunque l'attribuzione intelligente / stupido a comportamenti precede e giustifica la stessa attribuzione ad agenti: sarebbe quantomeno poco chiaro il senso di sostenere di una persona che è intelligente ma che si comporta sempre in modo stupido.

Nella ricerca di un significato condivisibile, si tratta di un demoltiplicatore di complessità rilevante: accettare che "essere intelligente" è solo un'abbreviazione per "comportarsi in modo generalmente intelligente". Qui l'intuizione si ancora più facilmente: un comportamento è intelligente se conduce a "leggere dentro" e a "leggere attraverso" ciò a cui il

del problema che stiamo esplorando qui: come scrisse Bridgman stesso, "il concetto è sinonimo del corrispondente insieme di operazioni".

³ www.britannica.com/technology/artificial-intelligence.

⁴ Nella letteratura dell'intelligenza artificiale il termine "agente", e quindi "agente intelligente" è spesso usato. Un agente è un'entità in grado di esibire un comportamento, e dunque di produrre risposte (*output*) sotto la sollecitazione di stimoli (*input*).

⁵ Si veda un'introduzione in it.wikipedia.org/wiki/Test_di_Turing.

comportamento si riferisce. Non che si tratti di una definizione chiara e univoca del concetto, ma è almeno una prima specificazione.

Nella forse inevitabile genericità di tutto ciò, riconosciamo inoltre che l'intelligenza è qualcosa non che si ha o non si ha, ma che si ha in gradi. O meglio, dunque, riconosciamo che ci si può comportare in modo più o meno intelligente. Ma allora se esseri umani possono comportarsi in modo più o meno intelligente, perché non dovrebbe poter essere lo stesso anche per altre specie animali? e per sistemi artificiali? Perché dunque un sistema artificiale non potrebbe avere un grano di intelligenza, e da quello eventualmente essere sviluppato (o svilupparsi, quando ne fosse in grado) nella sua intelligenza per gradi?

Con ciò, non abbiamo certo chiarito il concetto di intelligenza, né ci prefiggevamo di farlo, ma potremmo aver almeno suggerito la non fondatezza della posizione sciovinista secondo cui solo individui della specie *homo sapiens*, o comunque agenti non-artificiali, possono essere intelligenti. Piuttosto, una questione non ovvia è se esista uno "stile umano" di comportamento intelligente (nello stesso senso per cui potrebbe esistere uno "stile canino" di comportamento intelligente, e così via), e se poi ne esista uno solo o se invece l'intelligenza del comportamento sia riconosciuta diversamente, per esempio, in culture diverse. Questo pluralismo – la cui esplorazione è ben oltre le nostre competenze – renderebbe ancora più agevole l'accettazione di ulteriori "stili" di intelligenza, esibiti questa volta anche da agenti artificiali.

Intelligenza e comprensione

Sempre al fine di rendere conto della complessità del nostro tema, è il caso di accennare almeno a un'altra questione, che la proposta del test di Turing ha sollecitato: in breve, è possibile intelligenza *senza comprensione*? Oppure anche, assumendo che ciò che comprende è una mente: è possibile esibire intelligenza senza avere una mente, e dunque senza pensiero? Una volta che si sia accettato che intelligenti sono i comportamenti, la risposta è non problematicamente affermativa (e del resto è plausibile che chi sostiene l'esistenza di organizzazioni e processi intelligenti non implichi con ciò che essi hanno una mente), ma rimangono ampie divergenze quanto alla giustificazione di ciò.

La posizione meno impegnativa è quella della cosiddetta intelligenza artificiale "debole" (*weak AI*), che disaccoppia le questioni:⁶ non pretende che un agente che ha un comportamento intelligente debba comprendere ciò che sta facendo, ma nemmeno nega che intelligenza e comprensione possano essere diverse, dunque ammettendo, in pratica, che potrebbe esserci una differenza tra comportamento intelligente-senza-comprensione e comportamento intelligente-con-comprensione. In altri termini, chi sostiene questa posizione chiede solo di poter lavorare intorno all'intelligenza senza la necessità di un fondamento psicologico, qualsiasi cosa ciò significhi, e si accontenta di produrre cose utili perché in grado di esibire comportamenti intelligenti, qualsiasi cosa ciò significhi.

I fautori dell'intelligenza artificiale "forte" (*strong AI*) prendono invece una posizione esplicita: non c'è differenza tra intelligenza e comprensione, perché una mente non è altro che la funzione che produce un comportamento intelligente. Così che un sistema di intelligenza

⁶ Un ottimo testo di riferimento sull'intelligenza artificiale è il libro di S.J. Russell e P. Norvig, *Intelligenza artificiale – Un approccio moderno*, Pearson, Milano, in due volumi, 2010, 2005, che dedica un intero capitolo – di "Fondamenti filosofici" – all'analisi della distinzione tra intelligenza artificiale "debole" e forte".

artificiale in azione sta, letteralmente, pensando (nel suo articolo fondativo,⁷ Turing si chiedeva se le macchine possono pensare, e nel loro libro citato Russell e Norvig sintetizzano così: “l’asserzione che le macchine si comportino *come se* fossero intelligenti è chiamata “intelligenza artificiale debole”, e l’asserzione che le macchine che si comportano in questo modo stiano effettivamente pensando, e non stiano solo simulando di pensare, è chiamata ipotesi di “intelligenza artificiale forte”).

Come si può ben immaginare, questa posizione ha suscitato un dibattito molto ampio, in cui la visione contraria all’intelligenza artificiale “forte” è icasticamente rappresentata dall’argomento della “stanza cinese” proposto da John Searle,⁸ le cui conclusioni, senza entrare in dettagli, si possono sintetizzare così: mentre non ha senso distinguere un comportamento intelligente da una simulazione di comportamento intelligente, comprendere e simulare di comprendere sono condizioni diverse.

La questione è sottile: per fare un altro esempio, simulare di imparare è diverso da imparare? All’atto della sua nascita, sancito in una conferenza nel 1956, il programma di ricerca dell’intelligenza artificiale venne descritto come basato sulla “congettura per cui, in linea di principio, ogni aspetto dell’apprendimento o una qualsiasi altra caratteristica dell’intelligenza possano essere descritte così precisamente da poter costruire una macchina che le simuli”,⁹ dunque in una posizione coraggiosa in senso operativo ma cauta filosoficamente. Sardonico fu, al contrario, il commento che pochi anni dopo Edsger Dijkstra, una delle figure di riferimento della prima informatica, propose su questo tema: “i padri di questo settore erano piuttosto confusi: la questione se le macchine possono pensare è più o meno rilevante quanto la questione se i sottomarini possono nuotare”.¹⁰ E nuovamente Russell e Norvig chiosano: “la maggioranza dei ricercatori di intelligenza artificiale accetta l’ipotesi debole e non si preoccupa dell’ipotesi forte: se i loro programmi funzionano, non sono interessati se la si considera una simulazione di intelligenza o vera intelligenza”. Queste diatribe non negano la possibilità di un comportamento intelligente, e dunque la presenza di tali incertezze non ci impedisce di proseguire nella nostra esplorazione.

Intelligenza artificiale: un’entità complessa

Una volta accantonata l’obiezione di principio sull’impossibilità di comportamento intelligente da parte di un agente artificiale, e dunque una volta ammesso *che* un agente artificiale può comportarsi in modo intelligente, diventano sensati i due problemi successivi: *quando* un agente artificiale si comporta in modo intelligente e *a quali condizioni* un agente artificiale può farlo.

Non è sorprendente che intorno al primo problema – quando un agente artificiale si comporta in modo intelligente – le risposte non siano particolarmente interessanti, né convincenti: l’opzione à la Turing ricordata sopra non fa altro che trasferire il problema, lasciandolo però aperto data l’assenza di un modello di sfondo sul significato dell’intelligenza esibita per lo meno da esseri umani. Questa assenza pesa in particolare in riferimento all’intuizione che per essere intelligente un comportamento non dovrebbe essere

⁷ Tradotto nel libro curato da V. Somenzi, R. Cordeschi, La filosofia degli automi – Origini dell’intelligenza artificiale, Bollati Boringhieri, Torino, 1994.

⁸ Si veda un’introduzione in it.wikipedia.org/wiki/Stanza_cinese.

⁹ it.wikipedia.org/wiki/Conferenza_di_Dartmouth.

¹⁰ www.cs.utexas.edu/users/EWD/transcriptions/EWD08xx/EWD898.html.

“meccanico”. Secondo Douglas Hofstadter, “non appena si riesce a programmare qualche funzione mentale, immediatamente si smette di considerarla un ingrediente essenziale del “vero pensiero”. Il nucleo ineluttabile dell’intelligenza è sempre in quell’altra cosa che non si è ancora riusciti a programmare ... L’intelligenza artificiale è tutto ciò che non è ancora stato fatto.”.¹¹ Al di là dell’arguzia di questa considerazione, che presa alla lettera renderebbe in pratica vuoto l’insieme degli agenti artificiali intelligenti (come potrebbero, in un modo o nell’altro, non essere programmati?), è comunque vero che comportamenti che nel passato si ritenevano realizzati da sistemi necessariamente intelligenti – si pensi all’esempio del software per il riconoscimento della scrittura (*optical character recognition*, OCR) – oggi non sono più considerati tali. Potremmo interpretarlo come una sorta di effetto Flynn¹² in azione sugli agenti automatici: con il passare del tempo il loro quoziente di intelligenza aumenta, e l’intelligente di ieri potrebbe essere il mediocre di oggi.

Rimane in ogni caso che, come accade per gli esseri umani, anche per gli agenti artificiali non pare avere molto senso una distinzione netta tra comportamenti intelligenti e non. E’ vero che l’assenza di un criterio definito lascia aperta la porta agli effetti di hype cycle,¹³ ma forse è proprio nella natura del concetto di intelligenza di essere sfumato, così che potrebbe davvero essere accettabile considerare che un agente si comporti in modo intelligente quando sta realizzando un compito che nel contesto (prima di tutto storico, ma anche sociale) è giudicato difficile.

Più interessante è l’analisi del secondo problema: a quali condizioni un agente artificiale può comportarsi in modo intelligente? A questo proposito la vicenda storica dell’intelligenza artificiale è rilevante, e schematicamente la si può ricostruire così.

Nei primi trent’anni di sviluppo (indicativamente, 1960-1990), gli agenti artificiali sono stati progettati nell’ipotesi che si possa ottenere un comportamento intelligente da un sistema che opera secondo le leggi di una logica formale, a partire da fatti e regole forniti esplicitamente da esperti umani, attraverso un processo di elicitazione della conoscenza. Il caso più semplice è quello di una singola deduzione. Per esempio: se si osservano i sintomi *abc*, allora la diagnosi migliore è *xyz* (una regola); si osservano i sintomi *abc* (un fatto); dunque la diagnosi prodotta dal sistema è *xyz*. I termini impiegati per designare agenti di questo genere sono evocativi: “sistemi basati sulla conoscenza”, “sistemi basati su regole”, e più genericamente “sistemi esperti”. Per certi generi di problemi questa strategia è efficace, in particolare nelle situazioni in cui lo spazio delle soluzioni è noto e non troppo complesso, e sono esplicitabili le regole per avvicinarsi a una soluzione: un buon esempio sono i già citati sistemi di riconoscimento della scrittura (se è un’asta sufficientemente verticale, con un eventuale punto sovrapposto, è una “i”, e così via).

Ma ci sono numerosi problemi che gli esseri umani sanno risolvere, a volte anche senza con ciò ritenersi esperti, e la cui soluzione non sappiamo però descrivere attraverso un insieme di regole. Un esempio è quello del riconoscimento delle facce: siamo capaci di riconoscere persone dalle loro facce, ma non sappiamo descrivere come le riconosciamo. Quello che si può imparare dagli esperti (riconoscitori di facce, in questo caso) non è dunque un insieme

¹¹ D.R. Hofstadter, Gödel, Escher, Bach: un’Eterna Ghirlanda Brillante – Una fuga metaforica su menti e macchine nello spirito di Lewis Carroll, Adelphi, Milano, 1984.

¹² it.wikipedia.org/wiki/Effetto_Flynn.

¹³ it.wikipedia.org/wiki/Hype_cycle; un’interpretazione attuale nel caso dell’intelligenza artificiale è in www.gartner.com/smarterwithgartner/top-trends-on-the-gartner-hype-cycle-for-artificial-intelligence-2019.

di regole, ma l'associazione tra degli input (immagini di facce) e degli output (nomi di persone). Senza entrare in dettagli, le reti neurali artificiali funzionano su questo principio: se "addestrate" con un opportuno insieme di coppie (input, output), sono in grado di generalizzare quanto appreso, per esempio riconoscendo in un'immagine mai vista in precedenza la stessa faccia presente in un'immagine presente nell'insieme di addestramento, e quindi associando alla nuova immagine il nome della persona così identificata. La conoscenza, che nei sistemi esperti è dichiarata in un insieme di regole esplicite, nelle reti neurali è codificata nei valori dei parametri di configurazione della rete, assegnati automaticamente come risultato del processo di apprendimento. E perciò si può chiedere giustificazione a un sistema esperto di come ha ottenuto un certo risultato, e il sistema mostrerà l'elenco di regole (cioè la catena di deduzioni) che ha seguito per ottenerlo. Al contrario, tutto quello che una rete neurale può indicare è l'elenco dei valori dei suoi (tanti: migliaia ma spesso molti di più) parametri: non diversamente da quello che accadrebbe se cercassimo la spiegazione del parere di un esperto in un'immagine del suo cervello. Benché il nome "intelligenza artificiale" sia stato mantenuto, il cambiamento non potrebbe essere più radicale (e infatti l'intelligenza artificiale originaria è familiarmente chiamata GOF AI, "Good Old-Fashioned AI").

Scenari della nuova intelligenza artificiale

L'idea di rete neurale artificiale è precedente a quella di sistema esperto artificiale,¹⁴ ma è tornata alla ribalta solo negli ultimi dieci-quindici anni, nel contesto e come strumento abilitante fondamentale del *machine learning*.¹⁵ Anche scontando la delusione per le promesse relativamente non mantenute dai sistemi esperti,¹⁶ la plausibile ragione di questa transizione è duplice: da una parte, le reti neurali richiedono grande capacità di calcolo, oggi molto più disponibile di alcune decine di anni fa; dall'altra, le reti neurali funzionano generalmente tanto meglio quanto maggiore è la quantità (e la qualità) dei dati impiegati nel loro processo di apprendimento, e oggi, molto più che nel passato, i sistemi informatici sono connessi al mondo esterno, attraverso connessioni sociali (web, social media, ecc) e fisiche (il cosiddetto *Internet of Things*) che rendono disponibili appunto grandi quantità di dati.

Il ruolo dei dati è un'ulteriore differenza strutturale tra la "vecchia" e la "nuova" intelligenza artificiale: mentre un sistema esperto è un agente puramente simbolico, *disembodied*, una rete neurale può operare connessa "in tempo reale" a sensori e attuatori, come è per esempio necessario nel caso di un sistema di guida autonoma. Anche a questo proposito il parallelo con la struttura dei sistemi intelligenti naturali è suggestivo: "Il cervello è imprigionato all'interno del cranio, luogo silenzioso, buio e immobile: come può venire a

¹⁴ Come spiegato in it.wikipedia.org/wiki/Rete_neurale_artificiale, il modello di neurone artificiale di Warren McCulloch e Walter Pitts è del 1943.

¹⁵ Un buon esempio è TensorFlow (en.wikipedia.org/wiki/TensorFlow), un sistema software prodotto da Google per realizzare reti neurali e reso *open source* da qualche anno: è diventato lo standard di fatto.

¹⁶ Un caso evidente è quello dei sistemi di traduzione automatica, che appaiono avere prestazioni migliori quando, invece che essere fondati su conoscenze linguistiche esplicite, funzionano con reti neurali con apprendimento basato su giganteschi database di testi multilingue (per esempio, "Google Traduttore non applica regole grammaticali, poiché i suoi algoritmi si basano su analisi statistiche piuttosto che su analisi tradizionali basate su regole", it.wikipedia.org/wiki/Google_Traduttore), pur nell'evidente complessità del compito da svolgere.

sapere qual è l'aspetto del mondo esterno? ... Le uniche vie che dal mondo portino al cervello sono i fasci di nervi, come quelli che gli arrivano dagli occhi, dalle orecchie e dalla pelle. Come possono i segnali che giungono lungo questi nervi generare la nostra sensazione di "essere nel" mondo esterno? La risposta è che questa sensazione è una complicata illusione. In realtà noi non abbiamo mai un contatto diretto con il mondo esterno. Lavoriamo invece con modelli del mondo che ci costruiamo dentro il cervello".¹⁷

Siamo dunque sempre più circondati da agenti artificiali, più o meno intelligenti, la cui struttura mima quella degli agenti naturali intelligenti. Non è un caso che "cervello artificiale" fosse un termine evocativo 50 anni, e che oggi non lo si usi più, nella consapevolezza che in molte situazioni la disponibilità di un cervello, o un qualche sistema analogo, è una condizione necessaria ma non sufficiente di intelligenza.¹⁸

Il punto è però che la rete neurale di un agente naturale non può aumentare il numero dei suoi neuroni, è connessa al mondo esterno attraverso canali di trasmissione solo limitatamente migliorabili (per esempio con protesi come microscopi e altoparlanti), ed evolve con i tempi dell'evoluzione biologica. Le reti neurali artificiali non hanno nessuno di questi limiti, e in particolare possono essere connesse tra loro, per esempio in relazioni docente - discente interamente automatiche: cosa potranno diventare, e che nuove relazioni si stabiliranno (se mai una stabilità si raggiungerà...) tra agenti intelligenti naturali e artificiali?¹⁹

In uno scenario ottimistico, l'evoluzione degli agenti artificiali si svilupperà nella linea della complementarità con gli agenti naturali, verso una super-umanità simbiotica con agenti artificiali, in forme di "intelligenza collettiva aumentata" di cui già oggi troviamo le prime tracce nelle persone connesse alla rete (di persone e sistemi tecnologici) attraverso smartphone, una situazione che plausibilmente porterebbe una persona di solo cento anni fa a qualificarci come super-umani. In uno scenario pessimistico, l'evoluzione degli agenti artificiali si svilupperà nella linea della sostituzione degli agenti naturali.

¹⁷ Da M. Minsky, *La società della mente*, Adelphi, Milano, 1985, un libro che offre molti spunti di riflessione sugli agenti intelligenti come reti, "società", di agenti non-intelligenti.

¹⁸ Su temi come il "cervello nella vasca" è ricca di suggestioni l'antologia commentata di Douglas Hofstadter e Daniel Dennett, *L'io della mente – Fantasie e riflessioni sul sé e sull'anima*, Adelphi, Milano, 1985.

¹⁹ Per un approfondimento sul tema, si vedano Nick Bostrom, *Superintelligenza*, Bollati Boringhieri, Torino, 2018; Jerry Kaplan, *Intelligenza artificiale – Guida al futuro prossimo*, Luiss University Press, Roma, 2017; Melanie Mitchell, *Artificial Intelligence: A Guide for Thinking Humans*, Pelican, New York, 2019.