

Filosofia e Intelligenza Artificiale (IA): corso per università della Terza età 2020

Contenuti

Introduzione

- 1) Le origini filosofiche dell'IA
- 2) Nascita e storia dell'IA
- 3) Le realizzazioni più impressionanti dell'IA sotto il profilo filosofico
- 4) Come filosofia e IA interagiscono e si sviluppano a vicenda
- 5) I lati oscuri dell'IA
- 6) Le implicazioni etiche dell'IA

1° incontro

Introduzione

Stiamo vivendo la più radicale e sconvolgente delle rivoluzioni, una rivoluzione dagli sviluppi imprevedibili, che forse l'umanità non riuscirà nemmeno a gestire. Questa rivoluzione ha per base l'“Intelligenza Artificiale”.

L'Intelligenza Artificiale è ormai entrata diffusamente nelle nostre esistenze, anche negli ambiti di cui molti neppure sospettano.

La filosofia ha il dovere di comprenderla e governarla, anche perché le radici di questa rivoluzione affondano nella filosofia stessa, in particolare nell'esigenza di evitare gli errori dovuti alla nostra soggettività ed emotività nel pensiero e nella conoscenza.

Comprendere l'Intelligenza artificiale è altresì filosoficamente rilevante poiché essa non è una semplice riproduzione in qualche modo potenziata dell'intelligenza umana ma presenta alcune differenze significative, e non è del tutto perspicua. Infatti la grande stratificazione di ampie reti neurali, quella che consente i risultati più stupefacenti, non consente di capire come esse prendono le decisioni.

Ma forse sono proprio le differenze con la nostra intelligenza che ci stanno aiutando a mettere meglio a fuoco i problemi irrisolti della conoscenza umana, della mente, della coscienza e della scienza.

La definizione Il termine IA compare nel 1956 introdotto dal matematico John McCarthy (1927-2011) in occasione di una conferenza per distinguere il suo lavoro e quello dei suoi collaboratori dalla cibernetica, lo studio scientifico del controllo e della comunicazione negli animali e nelle macchine [Artificial intelligence Kaplan 2016 p.12].

Poco più tardi uno dei collaboratori di McCarthy, Marvin Minsky (1927-2016), considerato il padre dell'IA, formalizza il programma dell'IA in questi termini: **far fare alle macchine delle cose che richiederebbero intelligenza se fossero fatte dagli uomini**, ad es. giocare a scacchi, a GO e Poker,

scrivere novelle e articoli di giornale, diagnosticare tumori, decifrare le iscrizioni antiche, gestire l'economia, ecc..

In seguito le ricerche e le realizzazioni hanno "pluralizzato" l'Intelligenza artificiale cosicché per diversi è doveroso riconoscerne molte. In un recentissimo articolo sull'autorevole *Economist* leggiamo che "Intelligenza artificiale è un'espressione generica e imprecisa, che copre una gamma di tecniche che vanno dai sistemi di rispetto delle regole, sperimentati la prima volta negli anni cinquanta, all'apprendimento automatico basato sulla probabilità di oggi, nel quale i computer insegnano a se stessi a svolgere alcuni compiti."

Questa definizione riflette il punto di vista oggi prevalente anche se per alcuni è riduttiva poiché non è solo un insieme di tecniche, ma un "paradigma" o una "prospettiva" che ci consente di meglio comprendere, anche grazie alle numerose realizzazioni tecniche, diversi problemi – la conoscenza, la mente, la coscienza, ecc. - finora irrisolti.

Ritenere l'IA solo una tecnologia, volta a far fare alle macchine certe operazioni finora eseguite dagli umani, precisamente ciò che gli uomini fanno meglio, è la definizione di **IA debole**, cui si oppone l'**IA forte** - far sì che le macchine si comportino come uomini (test di Turing). Nel primo ambito rientrano la dimostrazione di teoremi, l'inferenza di leggi da dati particolari ("Bacon" il software che ha ricavato le leggi di Kepler dalle osservazioni raccolte da Tycho Brahe) giocare a scacchi, ecc. Pluribus in particolare, da un lato, battendo i campioni i poker, fa meglio di loro, e quindi apparterrebbe all'IA debole. Ma batte gli uomini comportandosi come loro, ossia bluffando.

1) Le origini filosofiche dell'IA

La formalizzazione degli antichi greci Possiamo far risalire le origini dell'IA all'antica Grecia allorché gli studiosi cominciarono a **formalizzare i loro procedimenti, ovvero a tradurli in sequenze manipolabili di segni rappresentativi senza doverne comprendere il significato**. Risultato della formalizzazione sono strutture puramente sintattiche manipolabili che ci consentono di conoscere la realtà. [Borzacchini "Il computer di Platone" 2005 pp.13-16].

Aristotele iniziò a formalizzare la logica, privilegiando il sillogismo o calcolo. Negli *Primi Analitici* all'inizio leggiamo: *La premessa A B sia dunque negativa universale. In tal caso, se A non appartiene a nessun B, neppure B apparterrà a nessun A. In effetti, se B appartenesse a qualche A, per esempio a C, non sarebbe che A non appartiene a nessun B: C infatti fa parte di B. D'altro canto, se A appartiene ad ogni B, anche B appartiene a qualche A. In effetti, se B non appartiene a nessun A, neppure A apparterrebbe a nessun B: si era supposto che A appartenga ad ogni B. **La dimostrazione è pure simile, quando la premessa è particolare**. In tal caso, se A appartiene a qualche B, è necessario che anche B appartenga a qualche A. In effetti, «se B non appartenesse a nessun A, neppure A apparterrebbe a nessun B. Per contro, se A non appartiene a B, non è necessario che anche B non appartenga a qualche A; ... [Organon vol.1 Aristotele a cura di Giorgio Colli pp.3-4]*

Il ragionamento procede senza che si conoscano i significati di A, B, C. Anche la relazione di appartenenza è formale. Si noti in particolare l'asserzione evidenziata in cui viene posta la similitudine tra dimostrazioni diverse, una con premessa universale, l'altra particolare.

Un altro esempio è costituito dall'uguaglianza nella matematica greca. Gli *Elementi* di Euclide segnano il passaggio da una geometria centrata sulla costruzione e sull'evidenza ad una geometria centrata sull'uguaglianza sintattica, ossia come fondamento dell'idea di quantità in quanto grandezza e non numero, e la similitudine. Non è cosa da poco se consideriamo che le dimostrazioni-costruzioni, direttamente visuali, sono compatibili solo con una matematica rigorosamente positiva e finita, che perciò non consente di trattare l'infinito. [Borzacchini "Il computer di Platone" 2005 pp.255 e seg.].

La formalizzazione è il presupposto dell'IA poiché consente ai computer di comprendere le istruzioni e realizzare i compiti ad essi assegnati. Infatti i linguaggi di programmazione hanno la peculiarità che non sono ambigui. Hanno una sintassi e delle regole ben precise che evitano possibili interpretazioni sbagliate da parte del computer. Per esempio, le tre istruzioni: numbers = [1,2,3,4]; Sum = sum (numbers); Print (Sum). [IL CONFINE DEL FUTURO Possiamo fidarci dell'intelligenza artificiale? 2019 di Francesca Rossi]

Hobbes. Leibniz e Cartesio Secondo il filosofo e neuroscienziato Daniel Dennett (1942- vivente), Hobbes (1588-1679) e Leibniz (1646-1716) nell'età moderna anticiparono l'IA. Essi tentarono infatti di esplorare le implicazioni dell'idea di suddividere la mente in piccole e puramente meccaniche operazioni [Brainchildren 1998 p.266]. Secondo questi due filosofi pensare non è altro che calcolare, ossia il pensiero non implica affatto quel potere singolarissimo e ineffabile che Cartesio gli aveva attribuito. Il pensiero umano non ha niente di speciale, tant'è che può essere automatizzato. Anzi occorre ridurlo a calcolo per liquidare le insuperabili ambiguità delle lingue umane. Leibniz a tal fine cercò di mettere a punto una *Caratteristica* (o scrittura) *universale*, arrivando tra i vari tentativi al linguaggio binario o "linguaggio macchina" (stringhe di 0 e 1 per esprimere qualsiasi pensiero umano).

Contemporaneo di Hobbes e padre della filosofia moderna, il celebre Cartesio (1596-1650), nel suo *Discorso sul metodo* anticipò il dibattito di fine 900 sull'IA con un esperimento immaginario di straordinaria attualità. Supponiamo di avere dinnanzi un automa che crediamo un uomo per l'aspetto e soprattutto per le sue risposte pertinenti alle nostre domande. Sembra che capisca quello che gli diciamo, ma in realtà per Cartesio non può comprendere il significato di ciò che dice. Per il filosofo francese l'automa, essendo privo della *res cogitans*, non può pensare, come invece fa chi comprende il significato di ciò che ascolta e dice, può solo simulare i discorsi umani senza comprenderne alcunché, come fa il pappagallo che ripete alcune parole o frasi ma senza comprenderne il senso. Cartesio considera infatti automi anche gli animali indistintamente, che invece noi ora sappiamo bene essere intelligenti. Gli studi negli ultimi decenni hanno concluso che in particolare primati e cetacei sono molti intelligenti. Cartesio sbaglia a causa del pregiudizio che fa dell'uomo la sola creatura per scelta divina capace di autentica intelligenza, ossia di comprensione del significato.

Russell Il programma di ricerca che possiamo chiamare “paradigma sintattico” avviato dagli antichi greci e proseguito nel 600 specie da Leibniz, fu ripreso dalla fine dell’800 da alcuni pensatori come Frege, Russell, Carnap, ecc.

Russell (1872-1970) proseguì l’opera di Frege perfezionando l’apparato logico di questi (uso estensivo del concetto di funzione e quantificatori), grazie a cui giunge ad elaborare la teoria delle descrizioni per liberare l’ontologia da oggetti non esistenti - ad es. le montagne d’oro - o addirittura impossibili - il quadrato-rotondo - che Meinong aveva ritenuto necessario postulare. Secondo Meinong siamo costretti a riconoscere come esistenti tali oggetti per il fatto che possiamo formulare proposizioni intorno ad essi (ad es. “Il quadrato-rotondo è inesistente” è una proposizione che sembra ineccepibilmente vera, ma proprio per concludere che essa è vera non possiamo fare a meno di pensare “il quadrato-rotondo” e quindi ammetterne l’esistenza, almeno come oggetto di pensiero, come oggetto logico, platonico). Per Russell possiamo analizzare le proposizioni senza dover postulare tali oggetti. Gli enunciati che sembrano riferirsi ai non esistenti non hanno una genuina funzione designativa, bensì solo una funzione predicativa, non sono nomi, sono solo descrizioni che devono essere analizzate attraverso quantificatori – (x) per ‘data una cosa qualsiasi’ o ‘per tutte le cose’; $(\exists x)$ per ‘data almeno una cosa’, che rappresentano i soggetti. Così è agevole negare degli oggetti senza doverli ammettere, ad es. “I quadrati-rotondi non esistono” equivale logicamente alla formula $(x) [-(Qx \cdot Rx)]$, ossia ‘nessuna cosa è quadrata e rotonda insieme’, o ‘per tutte le cose non si danno insieme le proprietà dell’essere quadrata e rotonda’.

In questo modo scopriamo la struttura logica nascosta degli enunciati, che non è facile individuare in quanto è spesso mascherata dalla loro forma grammaticale, e quindi realizziamo un modello con cui esaminare tutte gli altri enunciati ed eventualmente rigettarli se non hanno una struttura logica conforme a quella che è stata scoperta grazie alle ricerche volte a fare chiarezza sui principi della matematica. In altri termini, per Russell l’analisi è la sostituzione delle espressioni vaghe, equivoche e opache del linguaggio ordinario con notazioni, equivalenti nel significato, ma ricavate dall’apparato concettuale rigoroso della logica formale. Pertanto l’analisi è formalizzazione, creazione di un linguaggio trasparente che, a differenza di quello ordinario, colga la struttura della realtà.