

La filosofia contemporanea e l'I.A.

Piergiorgio Odifreddi

Novembre 1994

L'Intelligenza Artificiale è stata spesso attaccata con argomentazioni scientifiche, che si appellano a particolari risultati o esperimenti di pensiero, quali il teorema di Gödel (John Lucas), la camera cinese (John Searle), o la gravitazione quantistica (Roger Penrose). Le discussioni fondate su tali basi sono però spesso degenerare in dispute senza fine, sempre più dettagliate e sempre meno convincenti.¹

Una diversa linea di attacco, basata su un'argomentazione culturale, è stata iniziata da Hubert Dreyfus in *Quel che i calcolatori non possono fare* (1972), e completata da Terry Winograd e Fernando Flores in *Calcolatori e conoscenza* (1986). In sintesi, essi hanno semplicemente fatto notare che *il progetto dell'Intelligenza Artificiale è in contrasto con un'intera cultura filosofica del secolo XX.*

Il vantaggio di questa 'critica della ragione artificiale' (come Dreyfus sottotitola il suo libro) è evidente: non si tratta di discutere su argomenti specifici, ma solo di constatare un'incompatibilità culturale con una parte sostanziosa del pensiero contemporaneo. Lo svantaggio è altrettanto evidente: poichè la forza di un tale argomento contro l'Intelligenza Artificiale è soltanto proporzionale alla fiducia riposta in quel pensiero, in un caso limite questo argomento potrebbe anche lasciare completamente indifferenti.

⁰Testo di un intervento al Cenacolo Interdipartimentale di Torino, 2 Dicembre 1994, e al Circolo della Cultura e delle Arti di Trieste, 13 Dicembre 1994.

¹Si vedano ad esempio le decine di interventi in *Behavioural and brain sciences* sugli argomenti di Searle (3, 1980, pp. 417-457 e 13, 1990, 585-642) e Penrose (13, 1990, pp. 643-705).

Assunzioni epistemologiche

La contrapposizione fra Intelligenza Artificiale e pensiero filosofico contemporaneo ha radici antiche, e si può far risalire al dibattito su essere e divenire.

La filosofia del divenire, basata sull'evidenza percettiva dell'instabilità e del mutare di tutte le cose, si presentò come la risposta più immediata al problema epistemologico. Essa si espresse attraverso Eraclito nel VI secolo a.C. (“non ci si bagna mai due volte nello stesso fiume”), e Protagora nel V (“l'uomo è la misura di tutte le cose”).

La filosofia dell'essere, basata invece sulla stabilità di qualcosa al di là delle apparenze, fu proposta da Parmenide nel secolo V. Attraverso Platone ed Aristotele essa divenne egemone nella filosofia occidentale, e tale rimase fino al moderno ritorno ad Eraclito (a partire da Hegel).

Da un punto di vista matematico, le intuizioni di Platone ed Aristotele sono state realizzate in due forme complementari: da un lato, nella *teoria degli insiemi* il mondo viene analizzato in oggetti che possiedono proprietà esprimibili mediante predicati in un linguaggio formale; dall'altro lato, nella *logica classica* vengono descritti gli assiomi e le regole del ragionamento basato su tali predicati (in particolare, i principi di *non contraddizione* e del *terzo escluso*, ovviamente falsi per la filosofia del divenire).

La sistematizzazione di tale apparato matematico raggiunse forma compiuta tecnicamente nella *Ideografia* (1879) e nei *Principi dell'aritmetica* (1893) di Gottlob Frege (1848–1925), e letterariamente nel *Tractatus* (1921) di Ludwig Wittgenstein (1889–1951).

In queste opere si delinea l'epistemologia che sta alla base del progetto dell'Intelligenza Artificiale, e che si può riassumere come segue: *la struttura del linguaggio riflette la struttura del pensiero da un lato e quella del mondo dall'altro, ed è di natura formale*. In particolare, sia l'attività mentale che la conoscenza del mondo sono riducibili ad una serie di fatti e regole esprimibili in un linguaggio formale, e si possono dunque simulare su di un computer.

Sono proprio queste le assunzioni che vengono messe in crisi da una parte della filosofia contemporanea, con motivazioni e pensieri che ora passiamo a sintetizzare.

Husserl e gli oggetti

Edmund Husserl (1859–1938) propose, nelle *Ricerche logiche* (1900 e 1901) e nelle *Idee per una fenomenologia pura* (1913), una versione moderna della filosofia cartesiana, secondo cui la conoscenza trova fondamento nella esperienza immediata dei dati mentali individuali. L'idea è che ogni dato mentale debba però essere rielaborato, alla ricerca della informazione pura che esso contiene (detta *atto mentale*). La rielaborazione si effettua mediante una tecnica di ripulitura che consiste nell'escludere (*mettere fra parentesi*) tutto ciò che è in qualche modo elaborazione teorica, e presuppone quindi una sospensione di giudizio (*epoché*) nei confronti di ogni teoria. Husserl sintetizzò il suo approccio nel termine *fenomenologia*, che significa 'scoprire (*logos*) ciò che si mostra da sè (*phenomenon*)'.

Gli atti mentali ottenuti con l'analisi fenomenologica sono il mezzo attraverso cui il mondo ci si presenta nell'esperienza vissuta, mediante l'intuizione dell'essenza delle cose (i *noemi*). Nel tentativo di definire i noemi degli oggetti quotidiani, Husserl notò che la percezione è un fenomeno complesso e globale, che arriva alla parte attraverso il tutto. Da un lato, essa richiede infatti un *orizzonte esterno*, costituito dallo sfondo indeterminato senza il quale un oggetto non potrebbe ricevere attenzione.² Dall'altro, essa coinvolge un *orizzonte interno*, costituito dagli aspetti dell'oggetto che non sono considerati in quel particolare momento, pur essendo ugualmente presenti e determinanti.

Nel noema intervengono quindi aspetti sia esterni che interni dell'oggetto, alcuni dei quali fissi e altri variabili, che ricevono un'attenzione differenziata a seconda dell'occasione, e sono responsabili di attese e predeterminazioni collegate all'oggetto. In particolare, la definizione di un noema è un procedimento complesso, al punto che Husserl stesso finì per accorgersi che esso tendeva a divenire un 'compito infinito'.

La fenomenologia mise comunque seriamente in questione sul piano teorico la concezione di oggetti percepibili in modo indipendente dal contesto, e definibili staticamente mediante una serie di proprietà. Essa è stata suffragata sul piano pratico dalla psicologia sperimentale (in particolare, dal lavoro di Eleanor Rosch), che ha mostrato come in effetti gli oggetti vengano

²Il fenomeno fu anche scoperto dalla psicologia Gestalt negli stessi anni, sotto forma di contrapposizione *figura-sfondo*: si ricordi a questo proposito il *vaso di Rubin* (1915), in cui due profili di facce possono essere visti come il contorno di un vaso.

da noi classificati in base non ad insiemi che li contengono o a proprietà che li definiscono, ma ad immagini concrete e paradigmatiche a cui essi si avvicinano.

Dopo aver sentito suonare campane a morto sia per la teoria degli insiemi che per la epistemologia che su di essa si basa, l'Intelligenza Artificiale ha cercato di effettuare vari tentativi per integrare il concetto di noema: i *frames* di Marvin Minsky (1973), gli *scripts* di Roger Schanks (1975), i *prototypes* di Terry Winograd (1977). Essi si sono però arenati di fronte a difficoltà analoghe a quelle sperimentate da Husserl stesso, che in questo caso hanno assunto la forma di 'basi di dati illimitate'.

Husserl fu un erede del dualismo cartesiano: dopo aver isolato le essenze attraverso gli atti mentali, egli continuò infatti a ritenerle esistenti in modo indipendente ed oggettivo (anche se non necessariamente materiale). Ma nell'abbraccio con la fenomenologia il dualismo divenne inconsistente: da un lato, gli oggetti dipendono dal soggetto (che li 'crea' focalizzandosi su alcuni aspetti all'interno di una varietà potenzialmente infinita); dall'altro, gli oggetti sono indipendenti dal soggetto (che non crea la varietà di aspetti all'interno della quale si focalizza). Husserl fu dunque l'*ultimo* erede del cartesianesimo.

Heidegger e il mondo

Martin Heidegger (1889–1976) fu il primo a trascendere le basi del dualismo, che stanno nel porre da un lato la coscienza come soggetto che può (almeno in teoria) essere indipendente dal mondo, e dall'altro nel vedere il mondo come costituito da oggetti che esistono indipendentemente dal soggetto.

Il punto di partenza di *Essere e tempo* (1927) è infatti l'interdipendenza del mondo (*Essere*, o *Sein*) e delle entità che lo popolano (*esseri*). Poiché l'Essere e gli esseri non possono essere compresi indipendentemente l'uno dagli altri, la loro definizione dovrà dunque essere *ricorsiva*: ciò che ci interessa in questo contesto, la definizione degli esseri in funzione dell'Essere, è quindi soltanto metà della storia.³

Reificare il mondo vedendolo come costituito di oggetti è l'essenza della visione scientifica (la *tecnologia*): essa considera ogni cosa a completa dis-

³Ciò che invece interessava Heidegger era l'altra metà, e cioè la definizione dell'Essere in funzione degli esseri. Egli non riuscì però a scrivere la parte ad essa relativa di *Essere e tempo*, lasciando l'opera essenzialmente incompleta.

posizione dell'uomo, e riduce quindi l'Essere ad un solo essere. Tale visione presenta un duplice pericolo: da un lato, si disinteressa di tutto quanto non sia reificabile (classificandolo come *nulla*), nonostante esso comprenda aspetti fondamentali della vita dell'uomo quali l'arte, l'etica o la storia; dall'altro, finisce per ridurre l'uomo stesso ad oggetto, come dimostrano espressioni quali 'mercato del lavoro' o 'capitale umano'.

Fortunatamente, la tecnologia non è però l'unica possibilità per l'uomo: ne esiste anche una esistenziale (*existentiell*) o, meglio ancora, metaesistenziale⁴ (*existenzial*). Ponendosi nella tradizione metodologica di Husserl, Heidegger propose dunque una *fenomenologia dell'esistenza*: il soggetto acquista la sua identità attraverso l'appartenenza all'Essere (l'*esserci*, o *Dasein*), sperimenta questa sua appartenenza in modo contingente (l'*essere gettato*), e agisce in un contesto temporale (l'*essere-nel-mondo*).

Mediante il passaggio dall'antropocentrismo tecnologico all'umanesimo esistenziale si arriva così ad una visione più corretta dell'Essere (la *cura del mondo*⁵). In particolare, ci si può ora rivolgere alla vita quotidiana e notare che gli oggetti vi sono percepiti come parte di *processi* umani, da cui vengono isolati soltanto quando tali processi si intoppiano. Ad esempio: mentre inchiodiamo non ci accorgiamo di avere un martello e delle dita; ma quando qualcosa va storto e il martello finisce sulle dita, ci accorgiamo subito di entrambi. La reificazione degli esseri è dunque conseguenza del venire meno della loro più naturale e corretta visione funzionale (il loro *essere-pronti-per-l'uso*).

Anche la filosofia di Heidegger dunque, come già quella di Husserl, mette in crisi la visione logico-insiemistica di un mondo costituito da oggetti descrivibili mediante proprietà indipendenti dal contesto: da un lato, gli esseri sono funzioni dinamiche (culturali) e non oggetti statici (naturali); dall'altro,

⁴Il prefisso 'meta' va inteso non come in 'metafisica', ma come in 'metalinguaggio': 'metaesistenziale' significa dunque non 'al di là dell'esistenza', ma 'avente l'esistenza quale oggetto di analisi'.

⁵A volte le cure sono più mortali delle malattie: Heidegger entrò nel partito nazista nel 1933, quando Hitler salì al potere, e vi restò sino alla fine della guerra, dopo di che finì epurato.

Un legame esplicito tra la filosofia heideggeriana e il nazismo sta nel ruolo centrale assegnato alla morte, dalla prima in teoria e dal secondo in pratica. Poiché l'esserci (l'appartenenza dell'essere all'Essere) è temporale, esso non si completa che nel momento della sua fine, e può dunque raggiungere l'autenticità solo accettando e vivendo in anticipo la propria morte (*essere-per-la-morte*), con l'*angoscia* che conseguentemente ne deriva.

essi sono interdipendenti con l'intero Essere.

Una tale critica non è però puramente negativa: essa permette invece di innescare quello che Heidegger esplicitamente considerò come il vero movimento delle scienze, e cioè un processo di revisione dei fondamenti. In particolare, almeno due sistemi di rifondazione della matematica si possono considerare come tentativi di tener conto (in modo parziale) della sostituzione dei processi agli oggetti: il λ -calcolo e la *teoria delle categorie*. In entrambi gli enti basilari non sono più oggetti classificati secondo la relazione di appartenenza (ovvio retaggio, come il nome suggerisce, di una ideologia politica ben definita), ma funzioni da calcolare o comporre fra loro.

Nel campo informatico, l'opposizione fra visione oggettuale e funzionale è invece riflessa nelle cosiddette 'filosofie' DOS e *MacIntosh*. In particolare, l'approccio *user friendly* della seconda non è che una implementazione dell'intuizione heideggeriana dell'essere-pronti-per-l'uso, e il suo successo commerciale mostra quanto essa colga nel segno.

Merleau-Ponty e il corpo

Nella parte v del *Discorso sul metodo* (1637) René Descartes (1596–1650) sostenne che gli animali sono a tutti gli effetti delle macchine, mentre nessuna macchina può simulare le persone in modo così efficiente da superare quello che oggi viene detto il *test di Turing*.

In realtà, l'Intelligenza Artificiale si è arenata nella simulazione non tanto delle attività più umane ed 'elevate' quali il pensiero logico e matematico, ma proprio di quelle più animali e 'basse' quali il riconoscimento di immagini e suoni: sembrerebbe dunque che il problema della meccanizzazione di quello che Aristotele definì 'animale razionale' risieda più nel primo che nel secondo termine.

La filosofia di Maurice Merleau-Ponty (1908–1961), ne *La struttura del comportamento* (1942) e *La fenomenologia della percezione* (1945), fornisce una possibile spiegazione di questa apparente ironia della sorte. L'idea è che l'attività principale della coscienza, e cioè la intuizione dei noemi attraverso cui ci si presenta il mondo, è in realtà radicata nel corpo: esso è il nostro mezzo di avere un mondo, e la mente vi è incarnata.

Le varie tecniche corporee di interazione con l'ambiente si imparano con pratiche che possono anche essere parzialmente conscie e razionali, ma che il corpo trasforma in abilità fisiche da usare in maniera inconscia e intuitiva

da un lato, e integrate fra loro dall'altro. Ad esempio: si impara a guidare un'auto attraverso un'attenzione cosciente dell'articolazione di mani e piedi, ma la guida esperta richiede invece risposte immediate e riflessi. Ed i due livelli sembrano essere fra loro irriducibili: non c'è infatti alcuna evidenza che tutto ciò che è inconscio si possa rendere conscio, nè che ciò che viene razionalizzato a livello conscio fosse già tale a livello inconscio.

Il carattere fondamentale della distinzione fra conoscenza razionale (descrittiva) e corporea (prescrittiva) è anche dimostrato indirettamente dal fatto che ad essa si possono ricondurre alcune dicotomie memorabili e classiche: *esprit de géométrie e de finesse*, *spirito apollineo e dionisiaco*, *know-what e know-how*, *emisfero sinistro e destro*.⁶ La stessa informatica ha recepito la distinzione, sotto forma di *istruzioni dichiarative e procedurali*.

La mancanza del corpo (o di qualcosa di analogo ad esso) potrebbe dunque frapporre un ostacolo fondamentale alla simulazione della conoscenza umana, in particolar modo delle tecniche corporee, condannandola al livello primordiale del principiante da un lato, e alla mancanza di integrazione dall'altro.

Il secondo Wittgenstein e la società

Il linguaggio naturale ha un ruolo centrale nel programma dell'Intelligenza Artificiale: proprio nell'attività linguistica Descartes isolava infatti la differenza fra persone e animali, e il test di Turing si basa appunto sull'impossibilità di decidere se una conversazione sia effettuata da una persona o da una macchina.

La visione della teoria dei linguaggi formali come paradigma per la teoria dei linguaggi naturali, che sta alla base delle speranze di una loro simulazione meccanica, fu attaccata da Wittgenstein nelle *Ricerche filosofiche* (1953): un rifiuto tanto più significativo in quanto proveniente dall'interno stesso della tradizione analitica e positivista, che il *Tractatus* aveva contribuito a fondare.

Il secondo Wittgenstein abbandonò la rigidità del linguaggio formale, basato su definizioni che delimitano con chiarezza, in favore di *somiglianze di famiglia* che si sovrappongono e si incrociano. Nel suo nuovo approccio non si vede più il significato di una parola come un oggetto ad essa convenzionalmente associato, bensì come una famiglia di usi appresi e insegnati.

⁶Si vedano, rispettivamente: Blaise Pascal, *Pensieri*, 1670; Friedrich Nietzsche, *La nascita della tragedia*, 1872; John Dewey, *Natura e condotta dell'uomo*, 1922; e le ricerche di Roger Sperry, che gli valsero il premio Nobel per la medicina nel 1981.

L'apprendimento non può avvenire in base a regole (pena il regresso infinito richiesto dall'apprendimento delle regole stesse), ed è invece una prassi che si acquista padroneggiando una tecnica, sulla base del comportamento collettivo della società. Tale comportamento stabilisce che cosa sia corretto, definendo implicitamente una *forma di vita* umana. In particolare, non può esistere un *linguaggio privato*, perchè esso sarebbe incomunicabile, e non potrebbe quindi essere nè imparato nè compreso.

Nella visione di Wittgenstein il linguaggio naturale è dunque l'esatto opposto del linguaggio formale, e non può essere descritto in modo preciso nè, tanto meno, isolatamente dall'intero contesto sociale che lo genera e lo determina.

Conclusione

Sappiamo bene che, come una rondine non fa primavera, quattro filosofi non costituiscono una cultura. Il fatto è che essi sono soltanto rappresentanti di tendenze di pensiero che li trascendono, e conclusioni analoghe si sarebbero potute trarre citando al loro posto altri filosofi, quali Jacques Derrida, Michel Foucault, Hans Georg Gadamer, Jürgen Habermas, José Ortega y Gasset, Jean Paul Sartre, John Searle, . . . oltre che scienziati quali Gregory Bateson, David Bohm, John Eccles (premio Nobel per la medicina nel 1963), Humberto Maturana, Ilya Prigogine (premio Nobel per la chimica nel 1977), Roger Sperry (premio Nobel per la medicina nel 1981), Francisco Varela, e così via.

Possiamo allora dire che è tutta una cultura contemporanea a mostrare come la percezione ed il linguaggio coinvolgano il corpo, il mondo e la società interi: il progetto dell'Intelligenza Artificiale, che equivale all'impresa titanica di ridurre tutto ciò ad una serie di fatti e regole formali, sembra dunque destinato a far appunto la fine del *Titanic* (che richiese anni di progettazione e costruzione, e si inabissò nel suo primo viaggio).⁷ Corpo, mondo e società intervengono inoltre nella percezione e nel linguaggio in maniera olografica: poichè ogni parte coinvolge e presuppone il tutto, non possono esistere successi intermedi, e le formalizzazioni esistenti non sono dunque

⁷Un *Titanic* è effettivamente in cantiere: il CYC (da 'encyclopedia') di Doug Lenat, iniziato nel 1984 con un finanziamento di 50 milioni di dollari, e il cui scopo è la costruzione di una banca dati del senso comune di un occidentale del secolo XX, della grandezza prevista di un centinaio di milioni di asserzioni (*un* milione delle quali è stata inserita nei primi sei anni).

altro che distrazioni tecnologiche.

Oltre ad una funzione negativa di critica, la filosofia ha però anche spesso avuto quella positiva di proporre in forma letteraria direzioni che il pensiero matematico recepisce e precisa in forma tecnica. Stiamo pensando qui alla *matematica della filosofia*, esemplificata dai rapporti fra Platone e la teoria degli insiemi, Aristotele e la logica del prim'ordine, Hobbes e la teoria dei giochi, Hume e la logica induttiva, Leibniz e l'analisi non standard, Kant ed i teoremi di incompletezza, Hegel e la logica paraconsistente, . . . Metaforicamente, potremmo dire che nella storia intellettuale ogni cosa accade due volte: prima come filosofia, poi come matematica.

Le intuizioni della filosofia contemporanea possono quindi fornire lo spunto da un lato per una evoluzione della logica in direzioni che riflettano meno ingenuamente la complessità del pensiero umano (come proposto da Gian Carlo Rota⁸), e dall'altro per uno sviluppo dei computers rivolto non al rimpiazzamento dell'uomo, ma al soddisfacimento dei suoi bisogni e al raggiungimento dei suoi scopi (come proposto da Winograd e Flores).

Il compito non è impossibile (alcune delle istanze, da Husserl a Wittgenstein, sono già state recepite sia da intuizionismo e bourbakismo che dalla Intelligenza Artificiale⁹), nè dissennato (sia Kurt Gödel che Hermann Weyl furono profondamente influenzati dalle concezioni di Husserl¹⁰). Continuando in questa direzione, logica e informatica potrebbero tardivamente allinearsi alla cultura di un secolo che ormai volge al termine, e prepararsi ad affrontare il nuovo millennio con impostazioni e metodi meno datati di quelli ottocenteschi in cui invece ancora indulgono.

⁸Si vedano i capitoli XIV, XV e XXIV di Mark Kac, Gian Carlo Rota e Jacob Schwartz, *Pensieri discreti*, 1986 (??, 19??).

⁹Si veda ad esempio la raccolta di scritti curata da L. Steels, nel *Journal of Artificial Intelligence*, 1992.

¹⁰Si vedano Hao Wang, *Reflections on Kurt Gödel*, 1987, e Hermann Weyl, *Il continuo*, 1918 (Bibliopolis, 1977).